

**The exploitation of thermophiles and their enzymes for the construction of  
multistep enzyme reactions from characterised enzyme parts**

Submitted by William John Andrew Finnigan to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Biological Sciences

in September 2016

This thesis will be available for Library use, following publication of the material in chapters 4 and 5 and on approval by GSK. It will be available on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: .....

## Abstract

Biocatalysis is a field rapidly expanding to meet a demand for green and sustainable chemical processes. As the use of enzymes for synthetic chemistry becomes more common, the construction of multistep enzyme reactions is likely to become more prominent providing excellent cost and productivity benefits. However, the design and optimisation of multistep reactions can be challenging. An enzyme toolbox of well-characterised enzyme parts is critical for the design of novel multistep reactions. Furthermore, while whole-cell biocatalysis offers an excellent platform for multistep reactions, we are limited to the use of mesophilic host organisms such as *Escherichia coli*. The development of a thermophilic host organism would offer a powerful tool allowing whole-cell biocatalysis at elevated temperatures.

This study aimed to investigate the construction of a multistep enzyme reaction from well-characterised enzyme parts, consisting of an esterase, a carboxylic acid reductase and an alcohol dehydrogenase.

A novel thermostable esterase Af-Est2 was characterised both biochemically and structurally. The enzyme shows exceptional stability making it attractive for industrial biocatalysis, and features what is likely a structural or regulatory CoA molecule tightly bound near the active site.

Five carboxylic acid reductases (CARs) taken from across the known CAR family were thoroughly characterised. Kinetic analysis of these enzymes with various substrates shows they have a broad but similar substrate specificity and that electron rich acids are favoured. The characterisation of these CARs seeks to provide specifications for their use as a biocatalyst.

The use of isolated enzymes was investigated as an alternative to whole-cell biocatalysis for the multistep reaction. Additional enzymes for the regeneration of cofactors and removal of by-products were included, resulting in a seven enzyme reaction. Using characterised enzyme parts, a mechanistic mathematical model was constructed to aid in the understanding and optimisation of the reaction, demonstrating the power of this approach.

*Thermus thermophilus* was identified as a promising candidate for use as a thermophilic host organism for whole-cell biocatalysis. Synthetic biology parts



including a BioBricks vector, custom ribosome binding sites and characterised promoters were developed for this purpose. The expression of enzymes to complete the multistep enzyme reaction in *T. thermophilus* was successful, but native *T. thermophilus* enzymes prevented the biotransformation from being completed.

In summary, this work makes a number of contributions to the enzyme toolbox of well-characterised enzymes, and investigates their combination into a multistep enzyme reaction both *in vitro* and *in vivo* using a novel thermophilic host organism.

# List of Contents

Abstract	2
List of Contents	4
List of Figures	7
List of Tables	11
List of Supplementary Figures	13
Abbreviations	17
Symbols	21
Acknowledgements	22
Preface	23
<b>Chapter 1 - Introduction</b>	<b>25</b>
1.1 Biocatalysis	26
1.2 Thermophiles	46
1.3 Whole-cell biocatalysis at elevated temperatures	52
1.4 <i>In vitro</i> cascade reactions and the application of mathematical modelling	55
1.5 Carboxylic acid reductase enzymes and their use in the construction of enzymatic cascade reactions	60
1.6 Aims and objectives	70
<b>Chapter 2 - General Materials and Methods</b>	<b>73</b>
2.1 Materials	74
2.2 DNA methods	76
2.3 Protein purification and analysis	80
2.4 Other methods	83
<b>Chapter 3 - Structural and biochemical characterisation of <i>Archaeoglobus fulgidus</i> esterase reveals a bound CoA molecule in the vicinity of the active site.</b>	<b>85</b>
3.1 Authors	86
3.2 Preface for inclusion in this thesis	86
3.3 Abstract	87
3.4 Introduction	88

3.5	Results and Discussion	91
3.6	Methods	110
3.7	Acknowledgments and Other Information	118
<b>Chapter 4 - Characterization of carboxylic acid reductases as enzymes in the toolbox for synthetic chemistry</b>		<b>119</b>
4.1	Authors	120
4.2	Preface for inclusion in this thesis	121
4.3	Abstract	122
4.4	Introduction	123
4.5	Results	129
4.6	Discussion	147
4.7	Methods	157
4.8	Acknowledgements	161
4.9	Supporting Information	162
<b>Chapter 5 - Construction, modelling and optimisation of a seven enzyme, one-pot biotransformation</b>		<b>193</b>
5.1	Preface for inclusion in this thesis	194
5.2	Introduction	195
5.3	Results	202
5.4	Discussion	227
5.5	Methods	234
5.6	Supplementary Figures	240
<b>Chapter 6 - <i>Thermus thermophilus</i> as a host organism for whole-cell biocatalysis</b>		<b>269</b>
6.1	Introduction	270
6.2	Materials and methods	288
6.3	Results	295
6.4	Discussion	345
<b>Chapter 7 - General Discussion</b>		<b>357</b>
7.1	Multistep pathways in biocatalysis	358
7.2	Characterising enzymes	358

7.3	The characterisation of the thermostable esterase Af-Est2 for inclusion in the multistep reaction. _____	360
7.4	Characterisation of carboxylic acid reductase enzymes. _____	361
7.5	Completion of a multistep enzymatic reaction_____	362
7.6	The use of isolated enzymes with mathematical modelling and cofactor regeneration_____	363
7.7	Whole-cell biocatalysis at elevated temperatures. _____	364
7.8	Biocatalysis and synthetic biology in the future_____	366
7.9	Conclusions_____	366
<b>Chapter 8 - Bibliography_____</b>		<b>369</b>

# List of Figures

## Chapter 1

Figure 1-1 – Enzymes lower the Gibbs free energy of activation _____	28
Figure 1-2 - A reaction scheme for a simple one-substrate enzyme catalysed reaction_____	29
Figure 1-3 – The Michaelis-Menton equation _____	29
Figure 1-4 - Overview of the enzyme classes presented as oral or poster presentations at previous Biotrans Symposia. _____	32
Figure 1-5 - Whole-cell vs Isolated enzymes. _____	42
Figure 1-6 – Examples of the lipids of bacteria and archaea. _____	50
Figure 1-7 - CARs connect many enzyme reactions allowing a host of multi-step reactions to be constructed. _____	62
Figure 1-8 - Reaction scheme for an enzymatic cascade reaction for the hydrolysis and reduction of an ester through to an alcohol _____	63
Figure 1-9 - The reaction mechanism of $\alpha/\beta$ hydrolase enzymes. _____	65
Figure 1-10 - Mechanism of horse liver alcohol dehydrogenase. _____	69

## Chapter 3

Figure 3-1 – Effects of temperature and pH on AF-Est2 _____	95
Figure 3-2 – Inhibition of AF-Est2 _____	97
Figure 3-3 - A cartoon representation of the AF-Est2 structure. _____	98
Figure 3-4 - The interactions of the CoA ligand bound to AF-Est2. _____	102
Figure 3-5 - A comparison of the overall fold of AF-Est2 with that of related proteins. _____	107
Figure 3-6 - A comparison of the CoA binding to both AF-Est2 (cyan) and human carboxylesterase 1 (magenta) shown as cartoon models in the same orientation. _____	109

## Chapter 4

Figure 4-1 – Proposed mechanism of CAR enzymes. _____	127
Figure 4-2 - Phylogenetic tree of CAR enzymes. _____	132
Figure 4-3 – CAR activity for various benzoic acid derivatives, heterocycles and fatty acids. _____	137
Figure 4-4 – The activity of CAR enzymes in response to pH. _____	144
Figure 4-5 – The effects of temperature on CAR enzymes. _____	145
Figure 4-6 – Model for binding of substrates and inhibitors to the CAR enzyme. _	153

## Chapter 5

Figure 5-1 – Potential for utilising CARs in multi-step enzyme reactions _____	198
Figure 5-2 - A schematic of the seven enzyme reaction _____	203
Figure 5-3 - The operational window for temperature (A) and pH (B) for the seven enzyme reaction _____	205
Figure 5-4 - Validation of the esterase model. _____	210
Figure 5-5 - Validation of mpCAR and related enzyme models. _____	215
Figure 5-6 - Validation of ADH and ADH-PTDH modals _____	217
Figure 5-7 - Validations of combinations of the esterase, CAR and ADH reactions. _____	221
Figure 5-8 - Flow diagram for genetic algorithm used to optimise the reaction ____	223
Figure 5-9 - Validations of the optimised complete reaction _____	224
Figure 5-10 - Sensitivity analysis of the modelled un-optimised and optimised batch reaction _____	226

## Chapter 6

Figure 6-1 – A model of the DNA transporter of <i>T. thermophilus</i> _____	272
Figure 6-2 – Well defined parts are essential for the development of <i>T. thermophilus</i> as a host organism for whole-cell biocatalysis. _____	273
Figure 6-3 – A map of PslpA promoter and the beginning of the <i>slpA</i> gene. ____	280
Figure 6-4 – Standard BioBricks cloning method _____	285

Figure 6-5 – A <i>T. thermophilus</i> cell expressing AfEst2, mpCAR and ApADH, catalyzing the hydrolysis and subsequent reduction of methyl <i>p</i> -toluate through to <i>p</i> -tolyl alcohol. _____	287
Figure 6-6 – Template vectors used for the construction of both <i>T. thermophilus</i> BioBricks vectors _____	297
Figure 6-7 – <i>T. thermophilus</i> BioBricks plasmid 1 _____	298
Figure 6-8 – PCR amplification of two fragments for the construction of a second <i>T. thermophilus</i> BioBricks vector by Gibson assembly. _____	300
Figure 6-9 – <i>T. thermophilus</i> BioBricks plasmid 2 - pBBTTh _____	301
Figure 6-10 – pBBTTh functioning in both <i>E. coli</i> and <i>T. thermophilus</i> . _____	302
Figure 6-11 – Restriction digest of pBBTTh by <i>Xba</i> I and <i>Pst</i> I _____	303
Figure 6-12 - <i>Thermus thermophilus</i> HB27 codon usage _____	304
Figure 6-13 – The constructs used for measuring promoter strength _____	308
Figure 6-14 – SDS-PAGE analysis and western blot of total cell lysate of <i>T. thermophilus</i> expressing sfGFP and mRFP under the control of a range of promoters. _____	309
Figure 6-15 – Population average sfGFP fluorescence as a measure of promoter strength _____	310
Figure 6-16 – Constructs for the expression of AfEst2, ApADH and a complete synthetic operon. _____	312
Figure 6-17 - Growth curves of <i>T. thermophilus</i> constructs transformed with AfEst2 constructs (left), and their conversion of methyl <i>p</i> -toluate into <i>p</i> -toluic acid (right). _____	317
Figure 6-18 - Growth curves of <i>T. thermophilus</i> constructs transformed with ApADH constructs (left), and their conversion of <i>p</i> -tolualdehyde (right). _____	319
Figure 6-19 - SDS-PAGE and western blot on <i>T. thermophilus</i> total cell lysate, transformed with constructs for the expression of AfEst2 or ApADH only. ____	321
Figure 6-20 - Growth curves of <i>T. thermophilus</i> constructs transformed with synthetic operon constructs (left), and their conversion of methyl <i>p</i> -toluate (right). ____	323
Figure 6-21 - SDS-PAGE and Western blot of <i>T. thermophilus</i> total cell lysate, transformed with the constructs for the expression of a synthetic operon containing mpCAR, AfEst, ApADH and CtTran _____	325
Figure 6-22 – Oxidation of <i>p</i> -tolualdehyde by <i>T. thermophilus</i> at OD <sub>600nm</sub> 4.0 ____	327

Figure 6-23 – New constructs for the expression of codon optimized ApADH, ATN1-AND and ss- $\gamma$ -lactamase _____	331
Figure 6-24 - SDS-PAGE and western blot analysis of <i>T. thermophilus</i> total cell lysate, transformed with constructs for the expression of codon optimized ApADH, ATN1-ADH and ss- $\gamma$ -lactamase. _____	333
Figure 6-25 - Conversion of <i>p</i> -toluic acid or <i>p</i> -tolualdehyde by <i>T. thermophilus</i> cells expressing ApADH _____	335
Figure 6-26 – Conversion of <i>p</i> -toluamide by <i>T. thermophilus</i> cells expressing ss- $\gamma$ -lactamase. _____	336
Figure 6-27 - Thermostability of a number CAR proteins in comparison to ancCAR _____	338
Figure 6-28 – Constructs for the expression of codon optimized mpCAR, ancCAR and CtTran from the PslpA promoter _____	340
Figure 6-29 – SDS-PAGE and western blot analysis of <i>T. thermophilus</i> total cell lysate with soluble and insoluble fractions, transformed with constructs for the expression of codon optimized ancCAR, mpCAR and CtTran only, under the control of the PslpA promoter. _____	341
Figure 6-30 – Western blot on <i>T. thermophilus</i> total cell lysate, transformed with constructs for the expression of codon optimized ancCAR or mpCAR in combination CtTran (all FLAG tagged). And for expression of a Cas9 protein (FLAG tagged) from <i>S. thermophilus</i> . _____	343
Figure 6-31 – Hypothetical modular system for the creation of <i>T. thermophilus</i> vectors from established parts. _____	346
Figure 6-32 – A <i>lac</i> like operon on the pTT27 plasmid of <i>T. thermophilus</i> HB27 _	349
Figure 6-33 – The <i>pyrE</i> gene deletion method. _____	355



# List of Tables

## Chapter 1

Table 1-1 - Equations describing bi-substrate enzyme catalysed reactions	30
Table 1-2 – Six enzyme classes	31
Table 1-3 - The enzyme toolbox for synthetic chemistry (Turner and O'Reilly 2013; Pollard and Woodley 2007).	34
Table 1-4 - A comparison of whole-cell and isolated enzyme biocatalysis (Oroz-Guinea and García-Junceda 2013; Fessner 2015)	43
Table 1-5 - Adaptions that can increase thermostability (Singh et al. 2013).	52
Table 1-6 - Advantages and disadvantages of whole-cell biocatalysis at elevated temperatures (Frock and Kelly 2012).	53

## Chapter 2

Table 2-1 – Bacterial strains used in this study	74
Table 2-2 - Growth media used in this study.	75
Table 2-3 - Antibiotics used in this study.	75
Table 2-4 – Antibodies used in this study	76
Table 2-5 – PCR recipe for use with Phusion polymerase	77
Table 2-6 – PCR recipe for use with DreamTaq Green PCR Master Mix	78

## Chapter 3

Table 3-1 – The kinetic characterisation of AF-Est2 using pNP-esters with varying acyl chain length and pNP- benzoate as substrates.	92
Table 3-2 - The AF-Est2 data processing and structural refinement statistics.	116

## Chapter 4

Table 4-1– Carboxylic acid reductases chosen for this study.	131
--	-----

Table 4-2 – CAR activity against benzoic acid and its derivatives with electron donating and withdrawing groups_____	138
Table 4-3 – CAR activity against benzoic acid derivatives with the carboxylic acid group extended from the ring. _____	139
Table 4-4 – CAR activity against heterocycles and fatty acids _____	140
Table 4-5 – Half-life and degradation constant $K_D$ of CAR enzymes when incubated at 30 °C. _____	146

## Chapter 5

Table 5-1 – Kinetic equations _____	207
Table 5-2 – Kinetic parameters used in the model. _____	208

## Chapter 6

Table 6-1 – Selectable markers for <i>T. thermophilus</i> _____	275
Table 6-2 – Template and cloning vectors used in this study _____	289
Table 6-3 – DNA BioBrick parts used throughout this study. _____	289
Table 6-4 – Ribosome binding site strength _____	306
Table 6-5 – Percentage identity matrix comparing the AFO's from <i>P. furiosus</i> with those of <i>T. thermophilus</i> . _____	352

# List of Supplementary Figures

## Chapter 4

Supplementary Figure 4-1 - MUSCLE Alignment of adenylation domains of ANL superfamily members. _____	165
Supplementary Figure 4-2 - All CAR sequences used in this study to build phylogenetic tree in figure 2. _____	165
Supplementary Figure 4-3 – SDS-PAGE analysis of the purification of mpCAR, msCAR, niCAR, noCAR and tpCAR. _____	168
Supplementary Figure 4-4 – Carboxylic acid substrates _____	169
Supplementary Figure 4-5 – mpCAR kinetic data _____	170
Supplementary Figure 4-6 – noCAR kinetic data _____	173
Supplementary Figure 4-7 – tpCAR kinetic data _____	175
Supplementary Figure 4-8– niCAR kinetic data _____	177
Supplementary Figure 4-9 – msCAR kinetic data _____	179
Supplementary Figure 4-10 – Comparison of NADPH consumption with aldehyde production _____	180
Supplementary Figure 4-11 – Confirmation of benzaldehyde and <i>p</i> -tolualdehyde production by all the CARs tested. _____	183
Supplementary Figure 4-12 – Controls looking at the feasibility of using NADPH consumption to measure activity of CAR enzymes with nitro aromatic acid substrates. _____	184
Supplementary Figure 4-13 – Competitive inhibition of NADPH by NADP <sup>+</sup> _____	185
Supplementary Figure 4-14 – Competitive inhibition of ATP by AMP _____	186
Supplementary Figure 4-15 – Competitive inhibition of <i>p</i> -toluic acid by PPI _____	187
Supplementary Figure 4-16 – Mixed model inhibition of ATP acid by PPI _____	188
Supplementary Figure 4-17 – All 99 CAR sequences identified _____	189
Supplementary Figure 4-18 - Unrooted maximum likelihood phylogeny of the 124 CAR homologues. _____	191
Supplementary Figure 4-19 – Table of molecular weights and extinction coefficients calculated using ExPASy ProtParam. _____	191

## Chapter 5

Figure 5-1 - Supplementary Figure 1 - Differential equations for modelling _____	248
Figure 5-2 - Supplementary Figure 1 - SDS-PAGE analysis of the AF-Est2 protein purification _____	240
Figure 5-2 - Supplementary Figure 2 - SDS-PAGE analysis of the mpCAR protein purification. _____	241
Figure 5-2 - Supplementary Figure 3 - SDS-PAGE analysis of the ApADH protein purification. _____	242
Figure 5-2 - Supplementary Figure 4 - SDS-PAGE analysis of the PTDH protein purification. _____	243
Figure 5-2 - Supplementary Figure 5 - SDS-PAGE analysis of the ttPpiase protein purification. _____	244
Figure 5-2 - Supplementary Figure 6 - SDS-PAGE analysis of the tnPAP protein purification. _____	245
Figure 5-2 - Supplementary Figure 7 - SDS-PAGE analysis of the tnAK protein purification. _____	246
Figure 5-2 - Supplementary Figure 8 - SDS-PAGE analysis of the tePPK protein purification. _____	247
Figure 5-2 - Supplementary Figure 9 – Preliminary data on two PAP enzymes. _	249
Figure 5-3 - Supplementary Figure 1 – PTDH activity at pH _____	250
Figure 5-3 - Supplementary Figure 2 – PTDH thermostability and temperature vs activity _____	251
Figure 5-3 - Supplementary Figure 3 – tnPAP activity at pH _____	252
Figure 5-3 - Supplementary Figure 4 – tnPAP thermostability _____	253
Figure 5-3 - Supplementary Figure 5 – ttPPiase activity at pH _____	254
Figure 5-3 - Supplementary Figure 6 - ttPPiase thermostability _____	255

Table 5-2 - Supplementary Figure 1 - The kinetics of mpCAR with varying concentrations of ATP and p-toluic acid. _____	256
Table 5-2 - Supplementary Figure 2 – Parameters calculated fitting kinetics of mpCAR with varying concentrations of ATP and p-toluic acid to a sequential steady state equation. _____	256
Table 5-2 - Supplementary Figure 3 – The kinetics of mpCAR with varying concentrations NADPH. _____	257
Table 5-2 - Supplementary Figure 4 -Parameters calculated fitting kinetics of mpCAR with varying concentrations of NADPH to the Michaelis-Menten equation. ____	257
Table 5-2 - Supplementary Figure 5 - The kinetics of apADH in the reductive direction at 30 °C, pH 7.5 with varying concentrations of <i>p</i> -tolualdehyde. ____	258
Table 5-2 - Supplementary Figure 6 - Parameters calculated fitting kinetics of apADH in the reductive direction at 30 °C, pH 7.5, with varying concentrations of <i>p</i> -tolualdehyde to the Michaelis-Menten equation. _____	258
Table 5-2 - Supplementary Figure 7 - The kinetics of apADH in the reductive direction at 70 °C, pH 7.5 with varying concentrations of NADH. _____	259
Table 5-2 - Supplementary Figure 8 - Parameters calculated fitting kinetics of apADH in the reductive direction at 70 °C, pH 7.5, with varying concentrations of NADH, to the Michaelis-Menten equation. _____	259
Table 5-2 - Supplementary Figure 9 - The kinetics of apADH in the oxidative direction at 70 °C, pH 7.5 with varying concentrations of NAD <sup>+</sup> . _____	260
Table 5-2 - Supplementary Figure 10 - Parameters calculated fitting kinetics of apADH in the oxidative direction at 70 °C, pH 7.5, with varying concentrations of NAD <sup>+</sup> , to the Michaelis-Menten equation. _____	260
Table 5-2 - Supplementary Figure 11 - The kinetics of apADH in the oxidative direction at 70 °C, pH 7.5 with varying concentrations of <i>p</i> -tolyl alcohol. ____	261
Table 5-2 - Supplementary Figure 12 - Parameters calculated fitting kinetics of apADH in the oxidative direction at 70 °C, pH 7.5, with varying concentrations of NAD <sup>+</sup> , to the Michaelis-Menten equation. _____	261
Table 5-2 - Supplementary Figure 13 – The kinetics of PTDH with varying concentrations of NADP <sup>+</sup> and NAD <sup>+</sup> . _____	262
Table 5-2 - Supplementary Figure 14 - Parameters calculated fitting kinetics of PTDH with varying concentrations of NADP <sup>+</sup> and NAD <sup>+</sup> to the Michaelis-Menten equation. _____	262

Table 5-2 - Supplementary Figure 15 – Determining $k_{cat}$ for ttPPIase	263
Table 5-2 - Supplementary Figure 16 – The kinetics of tnPAP with varying concentrations of AMP	264
Table 5-2 - Supplementary Figure 17 - Parameters calculated fitting kinetics of tnPAP with varying concentrations of AMP to the Michaelis-Menten equation.	264
Table 5-2 - Supplementary Figure 18 - The kinetics of tnPAP with varying concentrations of Polyphosphate (PolyP)	265
Table 5-2 - Supplementary Figure 19 - Parameters calculated fitting kinetics of tnPAP with varying concentrations of polyphosphate to the Michaelis-Menten equation.	265
Table 5-2 - Supplementary Figure 20 - The kinetics of tnPAP with varying concentrations of MgCl	266
Table 5-2 - Supplementary Figure 21 – tnAK – CAR coupled assay to estimate $k_{cat}$	267
Figure 5-6 - Supplementary Figure 1 – Sensitivity analysis of apADH and apADH-PTDH modelled reactions.	268

## Abbreviations

[E <sub>T</sub> ]	Total enzyme concentration
5-FOA	5-fluoroorotic acid
ADH	Alcohol dehydrogenase
ADP	Adenine diphosphate
AFOs	Aldehyde ferredoxin oxidoreductase
AK	Adenylate kinase
Ala	Alanine
AMP	Adenine monophosphate
AORs	Aldehyde oxidoreductase
Asp	Aspartate
ATA	Transaminases
ATP	Adenine triphosphate
bp	base pair
BVMO	Baeyer-Villiger monooxygenases
CAI	Codon adaption index
CAR	Carboxylic acid reductase
CDS	Coding sequence
Chlr	Chloramphenicol
CK	Creatine kinase
CLEAs	Cross-linked enzyme aggregates
CLECs	Cross-linked enzyme crystals
CM	Cytoplasmic membrane
CoA	Coenzyme A
COOH group	Carboxylic acid group
Cys	Cysteine
Da	Dalton
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide
E factor	Environmental factor
EDTA	Ethylenediaminetetraacetic acid
ee	Enantiomeric excess
ERED	Enoate reductase
Est	Esterase
Glu	Glutamate
Gly	Glycine
GOI	Gene of interest
H-bonds	Hydrogen bonds
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HHDH	Halohydrin dehalogenase
His	Histidine

HLADH	Horse liver alcohol dehydrogenase
HPLC	High pressure liquid chromatography
I	Induced
iGEM	International genetically engineered machine (competition)
iso	all carbons except one form a continuous chain
$k_1$ and $k_{-1}$	Forward and reverse rate constants
Kan	Kanamycin
kbp	kilo base pair
$k_{cat}$	Turnover number
$K_M$	Michaelis constant
KRED	Ketoreductases
LB	Luria-Bertani
MAO	Monoamine oxidase
MCP	Tetrameric meta-cleavage product
MCS	Multiple cloning site
MES	2-( <i>N</i> -morpholino)ethanesulfonic acid
MOPS	3-( <i>N</i> -morpholino)propanesulfonic acid
mRNA	Messenger RNA
NAD(P)+	Nicotinamide adenine dinucleotide (phosphate)
NAD(P)H	Nicotinamide adenine dinucleotide (phosphate) reduced form
NI	Not Induced
NP	No promoter
NRPS	Nonribosomal peptide synthetases
NTP	Nucleoside triphosphate
OD	Optical density
OM	Outer membrane
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAP	polyphosphate-AMP phosphotransferase
PCR	Polymerase Chain Reaction
PDB	Protein data bank
PG	Peptidoglycan
Phe	Phenylalanine
PIPES	Piperazine- <i>N,N'</i> -bis(2-ethanesulfonic acid)
PK	pyruvate kinase
PMSF	Phenylmethylsulfonyl fluoride
pNP	<i>Para</i> -nitrophenyl
PolyP	Polyphosphate
PP <sub>i</sub>	Pyrophosphate
PPiase	Pyrophosphatase Enzyme
PPK	Polyphosphate kinase
PPT	phosphopantetheine
PTDH	Phosphite Dehydrogenase



RBS	Ribosome binding site
RFP	Red fluorescent protein
RNA	Ribonucleic acid
RPM	Revolutions per minute
rRNA	Ribosomal RNA
SCWPs	Secondary cell wall polymers
SD	Shine-Dalgarno
SDS	Sodium dodecyl sulfate
Ser	Serine
sfGFP	Super folder green fluorescent protein
slp	S layer protein
SOC	Super optimal broth with catabolite repression
ST	Total sensitivity indices
TAE buffer	Tris base, acetic acid and EDTA buffer
TB media	Thermus broth media
Tert	Tertiary
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
TTN	Total turnover number
UTR	Untranslated region
v/v	Volume by volume
Val	Valine
vmax	Maximum enzyme velocity
w/v	Weight over volume



## Symbols

$\alpha$	Alpha
Å	Ångström
$\beta$	Beta
°C	Degrees Celsius
$\delta$	Delta
\$	Dollar
$\gamma$	Gamma
$\Delta G^\ddagger$	Gibbs free energy of activation
$\mu$	Micro
%	Percent
$\pm$	Plus or Minus indicating error

## Acknowledgements

Thank you to:

All my supervisors: Nicholas Harmer, Jennifer Littlechild, Radka Snajdrova and Joseph Adams. Especially for all the time taken to attend meetings in Exeter or Stevenage, and their ideas, insight and advice throughout the project.

Particularly I would like to thank Nicholas Harmer for his support, mentoring and encouragement throughout the last four years.

The BBSRC and GSK for funding my project.

Halina Novak, Sam Barker and Tara Macey for being good friends when I needed some in a new city.

Callan Davies, Deniz Çizmecı and the associated crowd for some great times in the Lodge. Eyvallah!

My parents Timothy Finnigan and Susan Finnigan.

Everyone in the Biocatalysis building.

Everyone I met at GSK who were so welcoming during my three month placement there.

Chris Sayer for letting me team up with him for my first academic publication.

Adam Thomas successfully taking on the challenge of ancestral gene reconstruction of a CAR enzyme.

All my friends and family.

Anna Campbell for all the adventures of the past three years, and many more to come.

## **Preface**

Chapters 3, 4 and 5 of this thesis are reformatted manuscripts of papers which have been published, submitted or are intended for submission respectively. Each of these chapters includes a preface which explains the contributions of various authors to these works, and their publication status.

An extended introduction has been included as chapter 1, which introduces the areas these papers cover, and how they are related in the context of this thesis. However some repetition of this material may be present due to the need to introduce each paper in its own right.

A general materials and methods chapter has been included as chapter 2, to try and give more detail on the core techniques used throughout this thesis which may not have been covered in detail in the work intended for publication. Again some repetition of material is likely with the materials and methods sections of the individual chapters.

The manuscripts for these papers have been reformatted to match the formatting of this thesis, with figures renumbered accordingly. Also the bibliographies have been combined for a final bibliography at the end of the thesis.



# **Chapter 1 - Introduction**

## 1.1 Biocatalysis

### 1.1.1 Biocatalysis for green chemistry

In response to growing public concern and regulatory pressure, there is an ever increasing demand for the development of chemical manufacturing processes which “reduce or eliminate the use or generation of hazardous substances in the design, manufacture and application of chemical products”, which is the definition of green chemistry (Anastas and Warner, 1998). Sustainability and recycling are integral parts of the circular economy and biological methods are an excellent starting point in designing green processes (Banner *et al.*, 2011). Biocatalysis, the use of enzymes to carry out chemical reactions, offers a solution in meeting the green chemistry challenge. Recently increasing numbers of processes are employing biocatalysis, which ranges from the production of speciality to bulk chemicals (Wenda *et al.*, 2011). One of the greatest advantages enzymes have over their chemical counterparts is their often exquisite enantio- and regio- selectivity (Pollard and Woodley, 2007). This makes them particularly attractive for the production of pharmaceuticals and indeed this is where enzymes have seen the most use. In 2009, 72 % of drugs were chiral (not including protein drugs), a number that is likely to increase, therefore driving the uptake of biocatalysis further (Wells and Meyer, 2014).

As well as possessing excellent selectivity, enzymes are considered as a green alternative as they: I. can act as non-toxic catalysts, II. result in a decrease in waste production, III. are able to operate in mild reaction conditions such as at ambient temperature, pressure and pH, which can result in lower energy consumption, IV. work in aqueous media, V. may offer shorter synthesis routes, VI. prevent the large scale use of metals and organic solvents, and VII. are catalysts from a renewable feedstock (Wenda *et al.*, 2011).

The first biocatalysis reaction was carried out nearly a century ago when scientists realised that components of living cells could be used for chemical transformations (Bornscheuer *et al.*, 2012). More recent examples include the use of proteases in laundry detergents (Estell, Graycar and Wells, 1985), glucose isomerase to convert glucose into the sweeter tasting fructose (Bhosale, Rao and Deshpande, 1996) and



penicillin G acylase to make semisynthetic antibiotics (Bruggink, Roos and de Vroom, 1998). The first processes that were developed using enzymes found stability to be the major challenge. Enzyme immobilisation techniques were used to tackle this problem, also facilitating the reuse of the enzyme. Since then enzyme structures have allowed rational enzyme engineering for extended substrate scope and stability. More recently directed evolution approaches are rapidly modifying biocatalysts to meet industrial specifications (Bornscheuer *et al.*, 2012).

The field of biocatalysis is gaining increasing momentum with tools such as bioinformatics, rapid gene synthesis, access to enzyme structures and high throughput screening, which facilitate further development (Aldridge, 2013).

### 1.1.2 Enzymes

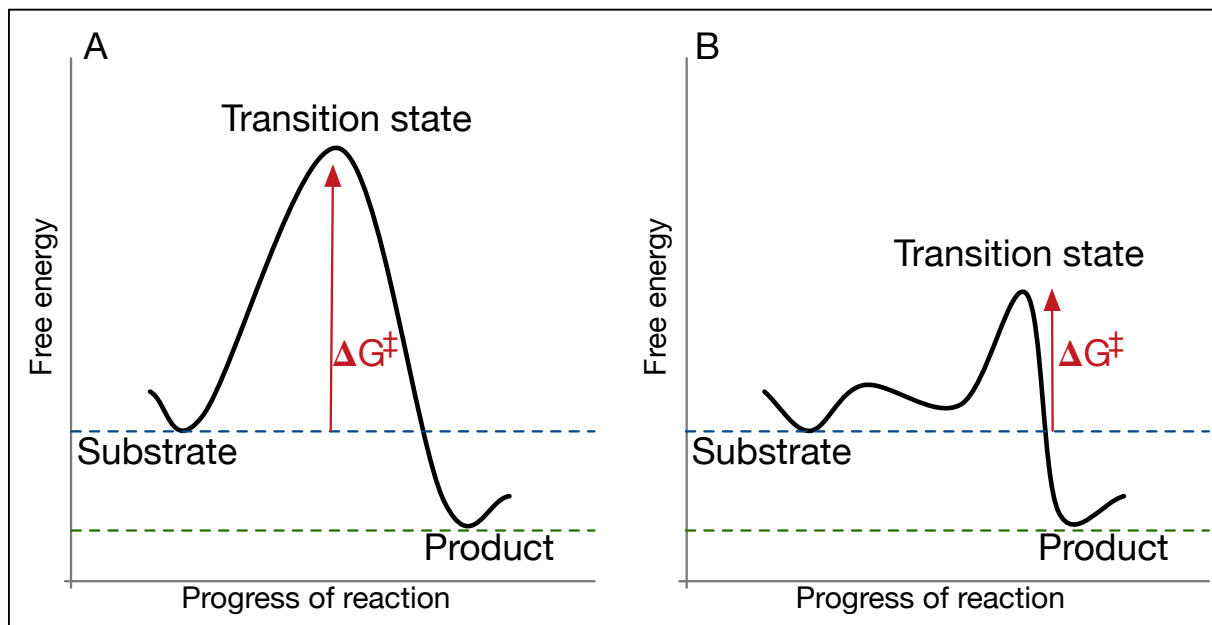
Enzymes are proteins that selectively catalyse biochemical reactions through the stabilisation of reaction transition states, offering a new reaction pathway for which the free energy of the transition state is lowered, decreasing the Gibbs free energy of activation ( $\Delta G^\ddagger$ ) (Figure 1-1) (Voet and Voet, 2004). The Gibbs free energy of activation is the difference in free energy between the transition state and the substrate (Stryer, 1995).

The first step in an enzyme catalysed reaction is the binding of substrates to the active site of an enzyme, forming an enzyme-substrate complex, with the substrates bound in a favourable orientation for a reaction to occur (Cook and Cleland, 2007). Active sites are generally formed of a cleft or crevice in the enzyme, and take up a relatively small part of the total volume of the enzyme. Amino-acid residues that come from different parts of the linear amino-acid chain come together in three-dimensions to form the active site. Much of the catalytic power of enzymes comes from their bringing together of substrates in favourable orientations in the enzyme-substrate complex (Stryer, 1995).

The substrate specificity of an enzyme is determined by the precise arrangements of atoms in the active site. The metaphor of a lock and key describing substrate binding was made in 1980 by Emil Fisher (Fischer, 1890). However, many enzymes have been shown to assume shapes complementary to their substrates only after

binding, leading to the induced fit model of interaction between substrate and enzyme (Koshland, 1995).

Like chemical catalysts, enzymes only accelerate reactions and cannot alter reaction equilibria. However, most reactions in biological systems do not occur at perceptible rates in the absence of enzymes. Enzymes are controlled in a number of ways including by feedback inhibition, by other regulatory proteins or by covalent modification such as phosphorylation (Stryer, 1995).



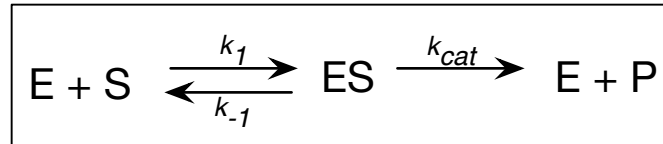
**Figure 1-1 – Enzymes lower the Gibbs free energy of activation**

Enzymes catalyse reactions by lowering the Gibbs free energy of activation. They offer a new reaction pathway whose transition state is lower, allowing a faster reaction to occur. Free-energy profiles of a reaction without enzyme (A), and with enzyme (B) is shown to demonstrate this. This figure is adapted from (Stryer, 1995).

### 1.1.3 Michaelis-Menten kinetics

For many enzymes, the rate of an enzyme catalysed reaction varies with substrate concentration. The classic equation describing the effect of changing substrate concentration on the velocity of an enzyme catalysed reaction is the Michaelis-Menten equation (Stryer, 1995). The use of the equation requires the use of the steady-state assumption, which is true for similar equations describing the effects of two or more substrates. The assumption states that the concentration of the

enzyme-substrate complex (ES in Figure 1-2) quickly reaches a steady state following the start of a reaction. It is also assumed that the substrate concentration (S) is greatly in excess of enzyme concentration (E), and that there is an absence of products (P). Using these assumptions the Michaelis-Menten equation (Figure 1-3) can be derived from the rate constants in Figure 1-2 (Cook and Cleland, 2007).



**Figure 1-2 - A reaction scheme for a simple one-substrate enzyme catalysed reaction**

The reaction scheme utilising first order rate constants for a one substrate enzyme reaction is shown in which the enzyme (E) and substrate (S) bind reversibly ( $k_1$  and  $k_{-1}$ ), forming an enzyme-substrate complex (ES). The enzyme catalysed reaction occurs and the product (P) is released. This is assumed to be the rate limiting step and is described by  $k_{cat}$ .

$$\begin{aligned} \text{A. } v &= V_{max} \cdot \frac{[S]}{K_M + [S]} \\ \text{B. } V_{max} &= k_{cat} \cdot [E_T] \\ \text{C. } K_M &= \frac{k_{-1} + k_{cat}}{k_1} \end{aligned}$$

**Figure 1-3 – The Michaelis-Menten equation**

- A. The Michaelis-Menten equation describing the rate of a reaction ( $v$ ) in terms of  $V_{max}$ ,  $K_M$ , and substrate concentration  $[S]$ .
- B.  $V_{max}$  is the maximum velocity that reaction can achieve if substrate concentration is extrapolated towards infinity. It is proportional to the  $k_{cat}$  multiplied by the total enzyme concentration  $[E_T]$ .
- C.  $K_M$  is often referred to as the Michaelis constant. It is an aggregate constant describing the rate of ES breakdown ( $k_{-1} + k_{cat}$  in Figure 1-2) divided by the rate of ES formation ( $k_1$  in Figure 1-2).  $K_M$  is described in units of substrate concentration and is equal to the substrate concentration which yields a rate ( $v$ ) of half  $V_{max}$ .

<b>Table 1-1 - Equations describing bi-substrate enzyme catalysed reactions</b>	
<b>Ordered sequential steady state and Random sequential</b>	$v = V_{max} \cdot \frac{[A] \cdot [B]}{(K_I^A \cdot K_M^B) + (K_M^A \cdot [B]) + (K_M^B \cdot [A]) + ([A] \cdot [B])}$
<b>Ordered sequential rapid equilibrium</b>	$v = V_{max} \cdot \frac{[A] \cdot [B]}{(K_I^A \cdot K_M^B) + (K_M^B \cdot [A]) + ([A] \cdot [B])}$
<b>Ping-Pong</b>	$v = V_{max} \cdot \frac{[A] \cdot [B]}{(K_M^A \cdot [B]) + (K_M^B \cdot [A]) + ([A] \cdot [B])}$
The three most commonly used rate equations for bi-substrate reactions. Substrate concentrations are shown as [A] and [B].	

Similar equations can be constructed to describe the effects of two substrates, although this also depends on the reaction mechanism (random sequential, ordered sequential or ping-pong), shown in Table 1-1. Sequential steady state, ordered sequential rapid equilibrium and ping-pong equations are most commonly used (Cook and Cleland, 2007). For three or more substrates, the method of King-Altman may be used to derive an equation (King and Altman, 1956).

### 1.1.5 The six classes of enzyme

All of the enzymes identified to date have been classified according to six Enzyme Commission numbers (EC numbers), according to the reactions they catalyse (Schomburg *et al.*, 2013). These are detailed in Table 1-2 (Meyer, Ghisalba and Leresche, 2014).

<b>Table 1-2 – Six enzyme classes</b>	
<b>EC number and name</b>	<b>Reaction catalysed</b>
<b>EC 1. Oxidoreductases</b>	Oxidation or reduction reactions. This involves the transfer of electrons from an electron donator to an electron acceptor, typically involving cofactors such as NAD(P)H or NAD(P) <sup>+</sup> , which must be regenerated for industrial use.
<b>EC 2. Transferases</b>	Transfer of a specific group from one substance to another, including methyl, acyl, amino, glycosyl or phosphate groups.
<b>EC 3. Hydrolases</b>	Hydrolysis of a substrate with water to form two products. This reaction can be made to work in the reverse direction when water is excluded from the reaction mixture.
<b>EC 4. Lyases</b>	Cleavage of various chemical bonds by means other than hydrolysis or oxidation. Can be made to work in the reverse direction.
<b>EC 5. Isomerases</b>	The conversion of one isomer of a molecule to another.
<b>EC 6. Ligases</b>	The linking of two molecules generally coupled with the hydrolysis of ATP or other NTP.

### 1.1.6 Enzyme discovery

Theoretically, almost any organic chemical reaction could be carried out by enzymes. However, we are limited by the enzyme activities known to us currently (Meyer *et al.*, 2013). Of the six enzyme classes, hydrolases have received by far the most attention in the past. Recently substantially more interest has been shown in oxidoreductases and to a lesser extent transferases, as shown in Figure 1-4 (Meyer *et al.*, 2013).

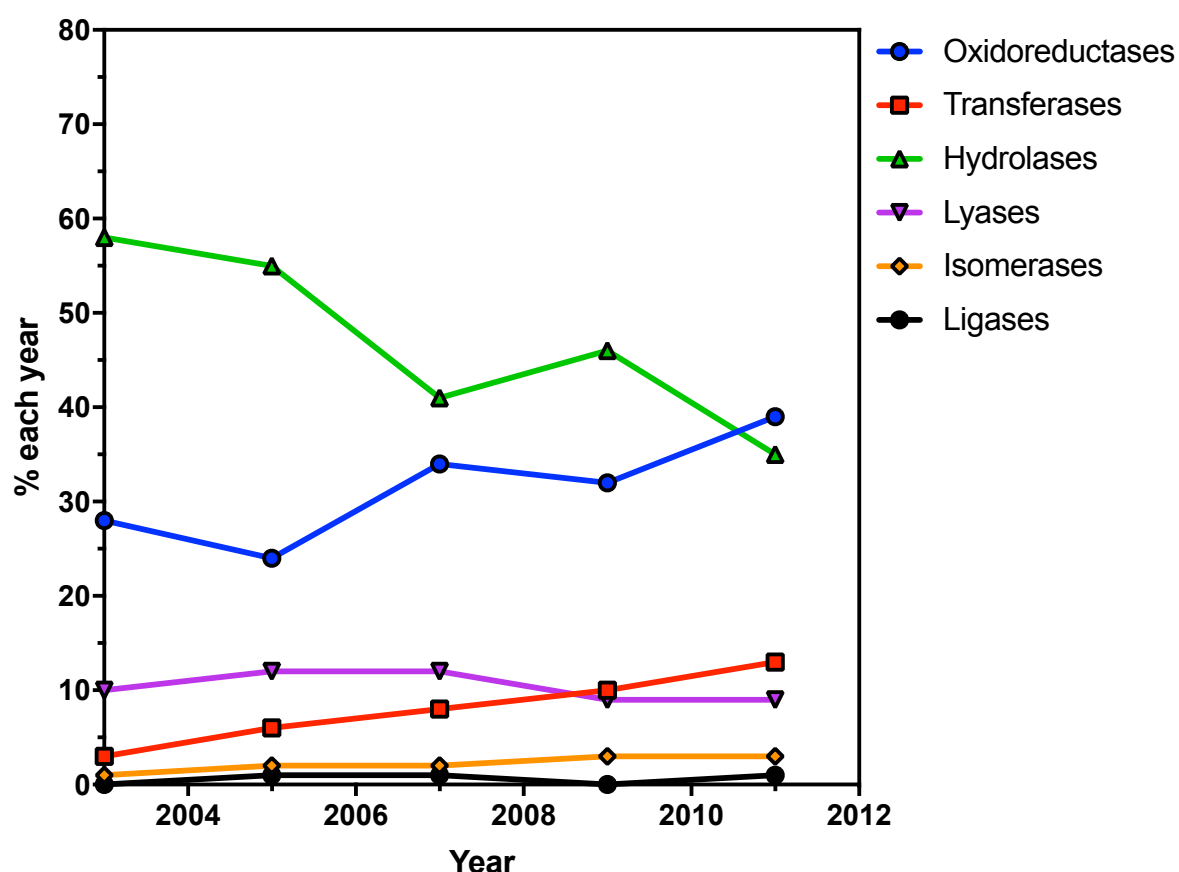


Figure 1-4 - Overview of the enzyme classes presented as oral or poster presentations at previous Biotrans Symposia.

Data are presented at a % of the total in each year – taken from (Meyer *et al.*, 2013).

### **1.1.7 The enzyme toolbox for synthetic chemistry**

Most industrial biotransformations are performed by chemists. For enzymes to be routinely used by chemists for the synthesis of new molecules, a toolbox full of different enzymes capable of carrying out a range of reactions and accepting a range of substrates is required (Meyer, Ghisalba and Leresche, 2014). Some enzymes in particular have secured their place in the enzyme toolbox for synthetic chemistry (Table 1-3), with many well-characterised enzymes available commercially, and a host of case studies demonstrating their uses (Woodley, 2008). Often a panel of enzymes of the same class is required to allow a broad range of substrates to be accepted, or to function over a range of conditions. Where a panel of well-characterised enzymes is available, it allows chemists to consider these reactions with more confidence when designing new synthetic routes to target molecules. Described in Table 1-3 are a number of enzymes considered part of the enzyme toolbox for synthetic chemistry, a toolbox which is expanding rapidly as sustainable chemistry becomes increasingly prevalent (Pollard and Woodley, 2007; Turner and O'Reilly, 2013).

**Table 1-3 - The enzyme toolbox for synthetic chemistry.**

(Pollard and Woodley, 2007; Turner and O'Reilly, 2013).

Enzyme class	Description	Reactions
Esterases, lipases, proteases, amidases and nitrilases	Hydrolases are probably the most widely used biocatalyst industrially and are especially attractive for the resolution of chiral compounds. Esterases and lipases catalyse the hydrolysis of ester bonds, with lipases catalysing the hydrolysis of longer acyl chain compounds such as fats and oils. Proteases or amidases catalyse the hydrolysis of amide bonds (Meyer, Ghisalba and Leresche, 2014). Nitrilase enzymes have also been gaining attention recently for the hydrolysis of nitriles into carboxylic acids and ammonia (DeSantis <i>et al.</i> , 2002). Hydrolase enzymes can also be made to work in the reverse direction when	$\text{R}^1\text{C}(=\text{O})\text{OR}^2 \xrightarrow[\text{Lipase}]{\text{Esterase}} \text{R}^1\text{C}(=\text{O})\text{OH} + \text{HO-R}^2$ $\text{R}^1\text{C}(=\text{O})\text{NHR}^2 \xrightarrow[\text{Protease}]{\text{Amidase}} \text{R}^1\text{C}(=\text{O})\text{OH} + \text{H}_2\text{N-R}^2$ $\text{R}^1\text{C}\equiv\text{N} \xrightarrow{\text{Nitrilase}} \text{R}^1\text{C}(=\text{O})\text{OH} + \text{NH}_3$



	they are used in the absence of water (Meyer, Ghisalba and Leresche, 2014).	
Halohydrin Dehalogenases (HHDHs)	Halohydrin dehalogenases (HHDHs) are lyases which can catalyse the ring closure of halohydrins producing epoxide rings, or the ring opening of epoxides via the addition of nucleophiles such as cyanide, azide or nitrite ions (Van Hylckama Vlieg <i>et al.</i> , 2001).	
Ketoreductases (KREDs) and Alcohol Dehydrogenases (ADHs)	Ketoreductases and alcohol dehydrogenase are enzymes that reversibly catalyse the reduction of aldehydes and ketones to primary or secondary alcohols. Various enzymes have been shown to accept a wide range of substrates including both aliphatic and cyclic compounds. They can also be used in the oxidative direction for the oxidation of alcohols to aldehyde and ketones, or even	

	aldehydes to carboxylic acids (Kaluzna, David Rozzell and Kambourakis, 2005). NAD(P) <sup>+</sup> or NAD(P)H are used as cofactors by these enzymes which must be regenerated for industrial use.	
Transaminases (ATAs)	Transaminases, also known as aminotransferases, catalyse the conversion of ketones to primary amines via the addition of an amine donor such as alanine, resulting in a ketone by-product. The reaction is dependent on pyridoxal 5'-phosphate (PLP) and is equilibrium driven. Transaminases allow the production of enantiomerically pure primary amines, which are valuable chiral intermediates for the preparation of pharmaceutical products (Green, Turner and O'Reilly, 2014).	$  \begin{array}{c}  \text{O} \\  \parallel \\  \text{R}^1 - \text{C} - \text{R}^2 \\  \text{Ketone starting material}  \end{array}  +  \begin{array}{c}  \text{NH}_2 \\    \\  \text{R}^3 - \text{C} - \text{R}^4 \\  \text{Amine donor}  \end{array}  \xrightleftharpoons{\text{ATA, PLP}}  \begin{array}{c}  \text{NH}_2 \\    \\  \text{R}^1 - \text{C} - \text{R}^2 \\  \text{Amine product}  \end{array}  +  \begin{array}{c}  \text{O} \\  \parallel \\  \text{R}^3 - \text{C} - \text{R}^4 \\  \text{Ketone by-product}  \end{array}  $
Carboxylic acid reductases (CARs)	Carboxylic acid reductases are a relatively new class of enzyme for biocatalysis which catalyses the reduction of carboxylic acids	$  \begin{array}{c}  \text{O} \\  \parallel \\  \text{R}^1 - \text{C} - \text{OH}  \end{array}  +  \begin{array}{c}  \text{NADPH} \\  \text{ATP}  \end{array}  \xrightarrow{\text{CAR}}  \begin{array}{c}  \text{O} \\  \parallel \\  \text{R}^1 - \text{C}  \end{array}  +  \begin{array}{c}  \text{PP}_i \\  \text{AMP} \\  \text{NADP}^+  \end{array}  $

	to aldehydes via an NADPH and ATP dependent reaction. The reduction of carboxylic acids chemically is difficult requiring harsh chemicals and carboxylic acid reductases represent a much greener alternative (Napora-Wijata, Strohmeier and Winkler, 2014).	
Enoate reductase (EREDs)	Enoate reductases, members of the “old yellow enzyme” family, catalyse the stereoselective reduction of alkenes featuring an electron withdrawing group. The use of enoate reductases in whole-cell reactions can be challenging due to undesired carbonyl reduction (Stuermer <i>et al.</i> , 2007).	$  \begin{array}{c} \text{R}^2 \\   \\ \text{R}^1 - \text{C} = \text{C} - \text{R}^3 \\   \\ \text{EWG} \end{array} + \text{NAD(P)H} \xrightarrow{\text{ERED}} \begin{array}{c} \text{R}^2 \\   \\ \text{R}^1 - \text{CH} - \text{CH} - \text{R}^3 \\   \quad   \\ \text{H} \quad \text{H} \\   \\ \text{EWG} \end{array} + \text{NAD(P)}^+  $

<p>Monoamine oxidase (MAOs)</p>	<p>Monoamine oxidases catalyse the oxidation of monoamines into imines. Typically the imine is hydrolysed by water to produce a carbonyl and ammonia, although a reducing agent can be used to reduce the imine which is useful for kinetic resolution of amines (Edmondson <i>et al.</i>, 2003).</p>	$  \begin{array}{c}  \text{R}^1\text{CH}(\text{NH}_2)\text{R}^2 \xrightleftharpoons{\text{MAO, O}_2} \text{R}^1\text{CH}=\text{NHR}^2 \xrightarrow{\text{H}_2\text{O}} \text{R}^1\text{C}(=\text{O})\text{R}^2 + \text{NH}_3 \\  \\  \text{R}^1\text{CH}(\text{NHR}^3)\text{R}^2 \xrightleftharpoons{\text{MAO, O}_2} \text{R}^1\text{CH}=\text{NR}^3  \end{array}  $
<p>Cytochrome P450 (P450 Monooxygenases)</p>	<p>Cytochrome P450's are heme-containing proteins which are generally the terminal oxidase enzymes in electron transfer chains, as well as having other key roles in primary and secondary metabolism or drug degradation. They are useful enzymes industrially for the oxy-functionalization of carbon-hydrogen bonds. However, they can also catalyse a range of reactions via a similar mechanism including, epoxidation, sulfoxidation and dealkylation, although</p>	$  \text{R-H} + \text{O}_2 + \text{NAD(P)H} \xrightarrow{\text{Cytochrome P450}} \text{R-OH} + \text{NAD(P)}^+ + \text{H}_2\text{O}  $

	hydroxylation is the most common (O'Reilly <i>et al.</i> , 2011; Urlacher and Girhard, 2012).	
Baeyer-Villiger Monooxygenases (BVMOs)	Baeyer-Villiger Monooxygenases (BVMOs) are flavin-containing, O <sub>2</sub> and NAD(P)H dependent enzymes. They catalyse the oxidation of ketone or heteroatom containing molecules for Baeyer-Villiger oxidations and sulfoxidations (Torres Pazmino, Dudek and Fraaije, 2010).	$\text{R}^2\text{CH}_2\text{C}(=\text{O})\text{R}^1 + \text{NAD(P)H} \xrightarrow{\text{BVMO}} \text{R}^1\text{C}(=\text{O})\text{OCH}_2\text{R}^2 + \text{NAD(P)}^+$ $\text{R}^1\text{S}(\text{R}^2)_2 + \text{NAD(P)H} \xrightarrow{\text{BVMO}} \text{R}^1\text{S}(=\text{O})(\text{R}^2)_2 + \text{NAD(P)}^+$
Aldolase	Aldolases catalyse asymmetric carbon-carbon bond formation between a nucleophilic carbonyl such as acetaldehyde, dihydroxyacetone or pyruvic acid, and a broad range of different electrophilic aldehydes (Turner and O'Reilly, 2013). They exhibit excellent specificity and recently attempts at tailoring substrate and stereospecificity <i>de novo</i> have been carried out (Windle <i>et al.</i> , 2014).	$\text{R}^1\text{CHO} + \text{R}^2\text{CH}_2\text{C}(=\text{O})\text{R}^3 \xrightleftharpoons{\text{Aldolase}} \text{R}^1\text{CH}(\text{OH})\text{CH}(\text{R}^2)\text{C}(=\text{O})\text{R}^3$

### 1.1.8 Expanding the enzyme toolbox

The total number of microbial species on Earth is estimated to be  $10^{12}$  (Locey and Lennon, 2016), offering a huge wealth of protein diversity (Guazzaroni, Silva-Rocha and Ward, 2015). Vast numbers of bacterial genomes are now being sequenced with third generation sequencing potentially able to produce a finished genome in a few hours (Land *et al.*, 2015). Bioinformatics tools such as the commonly used basic alignment search tool (BLAST) allow genomes to be mined for protein sequences of interest (Altschul *et al.*, 1990). However, whilst between  $10^3$ - $10^5$  microbial species are estimated to be present in just 1 g of soil, less than 1 % can be cultured using existing methodologies. Metagenomics avoids the need to isolate and cultivate individual microorganisms and has been proven as a powerful tool in identifying novel enzymes with potential for industrial applications (Guazzaroni, Silva-Rocha and Ward, 2015).

Natural enzymes are generally not optimised for industrial purposes. Biocatalysts with increased turnover numbers and lower  $K_M$  values for their industrial substrate are highly desirable. Increased catalytic efficiency on insoluble substrates, increased stability at both higher temperature at extremes of pH, higher tolerance to end product inhibition and stability in various solvents are all desirable attributes which can be engineered into an enzyme depending on the requirements (Dalby, 2007).

Two main approaches to enzyme engineering have been used previously. The first enzyme engineering attempts used rational design requiring detailed knowledge of the enzyme structure. Subsequently, random mutagenesis has been used to generate libraries of enzyme variants which are then screened in a process commonly referred to as directed evolution. Screening approaches are required to test all of the enzyme variants created, often with subsequent rounds of evolution towards a directed target. Currently, this is the most problematic part of a directed evolution process, with some types of reaction particularly difficult to screen for, especially when multiple traits are required. In recent years some library sizes have become much larger requiring high throughput screening technologies to be used (Woodley, 2013).

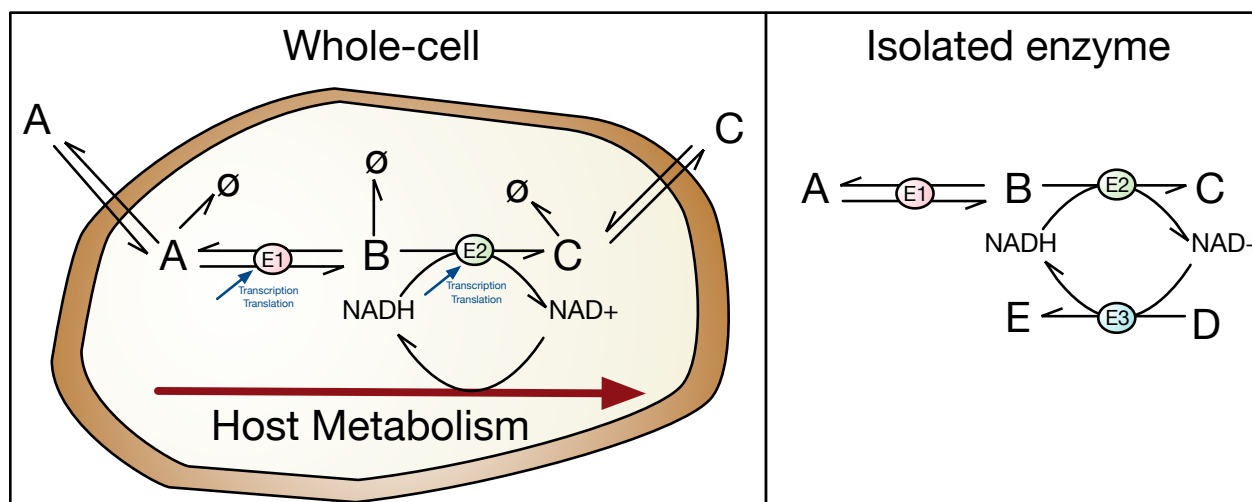
Structural knowledge, catalytic insight and multiple sequence alignment of protein sequence datasets can allow the construction of much smaller high-quality libraries with a greater chance of success, requiring fewer variants to be screened (Woodley, 2013). Recent trends have also seen the use of computational tools such as molecular modelling or even *de novo* design to complement structural analysis and directed evolution (Damborsky and Brezovsky, 2014).

### **1.1.9 Using enzymes - isolated enzymes or whole-cells**

Enzymes may be used as isolated enzymes or maintained within the host cells expressing the enzyme for use as a whole-cell biocatalyst, as shown in Figure 1-5. The use of enzymes in whole cell systems have a number of advantages such as including cofactor regeneration utilising host metabolic pathways, improved stability of the enzyme inside the cell and no requirement for isolation of the enzyme from the host resulting in a less complex upstream process (Ladkau, Schmid and Bühler, 2014). However, whilst whole-cell biocatalysis can offer an easier upstream process, it suffers from a greater complexity downstream as products must be isolated from cells once the reaction is complete, and unwanted side reactions can occur resulting in a less pure product (Duetz, Van Beilen and Witholt, 2001). Furthermore, achieving a high productivity can be challenging as diffusion across cell membranes is often rate limiting, and high substrate concentrations can be hard to achieve due to toxicity effects. Furthermore, the reaction itself is more difficult to control as any changes must be engineered into the microorganism (Schrewe *et al.*, 2013; Ladkau, Schmid and Bühler, 2014).

Isolated enzymes allow the downstream processing of a reaction to be much cheaper and simpler (du Preez *et al.*, 2015). Side reactions are reduced and a high purity product can be easily isolated. Possibly the biggest advantage is the ability to simply control these reactions. Changing the concentrations of enzyme, substrate, co-solvent, pH and temperature can allow the most efficient process possible to be designed (Fessner, 2015). However, the use of isolated enzymes also comes with a number of challenges and problems. Isolated enzymes can be less stable than they are inside the cell, and there is no possibility of producing more enzyme as there is in a whole-cell system. Therefore stability is an important factor for the use of isolated enzymes, with high stability necessary for an economical process. Cofactor

regeneration systems can also be necessary (Oroz-Guinea and García-Junceda, 2013). Furthermore, to properly optimise an *in vitro* reaction enzymes must be properly characterised (Vasić-Rački, Findrik and Vrsalović Presečki, 2011). A comparison of the advantages and disadvantages of whole-cell and isolated enzyme biocatalysis is shown in Table 1-4.



**Figure 1-5 - Whole-cell vs Isolated enzymes.**

A comparison of the use of whole-cells and isolated enzymes for biocatalysis. An example reaction is shown consisting of a reversible reaction catalysed by enzyme E1, and an oxidoreductase enzyme E2 using NADH. In this example NADH can be regenerated by the host metabolism using whole-cell biocatalysis, or can be regenerated using a third enzyme E3 if isolated enzymes are used. Whole-cell biocatalysis offers a simpler upstream process in that enzymes do not need to be purified, and cofactors can be regenerated. However side reactions can occur and diffusion across the cell membrane, as shown in the figure, can be rate limiting. The use of isolated enzymes overcomes some of these problems but the addition of a cofactor regeneration system is often needed, and enzymes must be isolated and purified before use.



<b>Table 1-4 - A comparison of whole-cell and isolated enzyme biocatalysis</b> (Oroz-Guinea and García-Junceda, 2013; Fessner, 2015)	
<b>Whole-cell advantages</b> <ul style="list-style-type: none"> <li>• No need to isolate enzymes.</li> <li>• Enzymes are generally more stable in the cellular environment.</li> <li>• Cofactors can be regenerated using the host metabolism.</li> </ul>	<b>Isolated enzymes advantages</b> <ul style="list-style-type: none"> <li>• Processes can be controlled and optimised simply by changing parameters such as enzyme or substrate concentration.</li> <li>• Faster rates as no cellular membrane.</li> <li>• Few side reactions occur.</li> <li>• High substrate concentrations and solvents are better tolerated.</li> <li>• Product purification easier.</li> </ul>
<b>Whole-cell disadvantages</b> <ul style="list-style-type: none"> <li>• Diffusion of the substrate or product across the cell membrane can be rate limiting.</li> <li>• Side reactions can occur resulting in lost substrate or product.</li> <li>• Cells may be less robust against high substrate concentrations or the use of solvents.</li> <li>• Process is difficult to control.</li> </ul>	<b>Isolated enzymes disadvantages</b> <ul style="list-style-type: none"> <li>• Cofactor regeneration is needed for some enzymes.</li> <li>• Enzymes may be less stable and are not regenerated.</li> <li>• Enzyme isolation is required.</li> </ul>

### 1.1.10 Are enzymes always green?

Whilst enzymes do have a number of attractive properties, they are often cited as a green alternative without an evaluation of the economic feasibility or environmental friendliness taking place. For example, many synthetically interesting substrates are hydrophobic, possessing poor solubility in aqueous media which is generally used for enzyme reactions (Domínguez de María and Hollmann, 2015). To overcome this, reactions must be carried out in the millimolar range resulting in large volumes of aqueous media being used. Once the reaction is complete this is considered waste which must be further treated, requiring energy and reagents to do so resulting in a poor environmental (E) factor (defined as the ratio of the mass of the waste per mass of product) overall (Leuchs, Na'amnieh and Greiner, 2013). Furthermore diluted biocatalysis with low substrate loadings is economically quite inefficient and can only be profitable if the product is of high value. As a rule of thumb for fine chemicals, processes should run at a minimum of  $> 50 - 100 \text{ g substrate L}^{-1}$  (Woodley, 2013).

To overcome these solubility challenges some reactions can be run with water-miscible co-solvents such as methanol, ethanol, *tert*-butanol or *iso*-propanol. Whilst this can allow higher substrate concentrations to be used it also results in higher product concentrations which can result in undesirable product inhibition (Stepankova, Damborsky and Chaloupkova, 2013). Hydrophobic co-solvents can also be used for a two liquid phase system, acting as both a substrate reservoir and product sink. Solvents such as butyl acetate, *tert*-butyl methyl ether or 2-methyl tetrahydrofuran have been used for this purpose (Woodley and Lilly, 1990). Other interesting approaches include the use of ionic liquids (Yang and Pan, 2005).

### 1.1.11 Industrial Biocatalysis

According to a recent review of biocatalysis by the Swiss Industrial Biocatalysis Consortium (SIBC), the industries using biotechnology in terms of value added can be ranked as 1. Pharmaceuticals, 2. Food industry, 3. Agrochemicals, 4. Cosmetics, 5. Polymers and 6. Others. While biocatalysis is most widely used in the production of pharmaceuticals, unfortunately timeline compressions during development mean that biocatalysis often represents the second generation process choice. Where enzymes are employed in the production of pharmaceuticals,

oxidoreductases followed by hydrolases are by far the most commonly used enzyme classes (Meyer *et al.*, 2013).

A number of enzymes were also identified by the consortium as being a priority in the development of new enzymes for biocatalysis. NAD(P)H dependent dehydrogenases for the reduction of ketones, ketoacids, olefins and other chemicals, and the oxidation of alcohols by alcohol dehydrogenases were identified as the top enzyme reactions which would ideally be developed (Meyer *et al.*, 2013).

In 2012 biotechnology as a whole generated at least \$324 billion in the US alone, with industrial products (biofuels, enzymes, biomaterials and biochemicals) representing at least \$104 billion of this. Biotechnology continues to grow more quickly than other industries although it is difficult to predict for how long this may continue (Carlson, 2016).

The opportunity for combining bio-bio or bio-chemo catalysis steps has been identified as having great potential, especially as multiple enzyme steps may be combined utilising the same reaction conditions. However commercial processes utilising more than one enzyme step are still rare (Wells and Meyer, 2014).

### **1.1.12 Multistep biocatalysis**

Reactions in which multiple catalysts work together for one-pot reactions are becoming increasingly popular. Combining reactions into one process removes the need to isolate substrates and products at each step, which gives excellent cost and productivity benefits (Turner and Truppo, 2013). Multistep pathways can also allow reversible processes to become irreversible, or remove inhibitory products (Oroz-Guinea and García-Junceda, 2013).

Enzymes are well suited to integration into multi-step pathways and indeed this is how they function in nature, catalysing complex metabolism networks (Ravasz *et al.*, 2002). In contrast chemical steps will often have differing reaction conditions making multi-step one pot reactions challenging to achieve (Wells and Meyer, 2014).

Enzymes are particularly attractive for the conversion of simple and cheap starting materials into high value enantiomerically pure products (Turner and Truppo, 2013).

Synthetic biology concepts offer an excellent framework from which to consider the development of multi-step biocatalysis pathways. Synthetic biology aims to allow the rational design of complex biological systems: if biocatalysis is to become a widely adopted technology for sustainable chemical synthesis this is something that should be embraced (Bujara *et al.*, 2010).

## 1.2 Thermophiles

### 1.2.1 Thermophiles in Biocatalysis

Enzyme stability can be a critical parameter in the development of effective biocatalytic processes (Bujara *et al.*, 2010). Enzymes from thermophiles offer an excellent starting point in overcoming this problem in that they have increased stability not only at high temperatures but also against organic solvents, extremes of pH or proteolytic enzymes (Atomi, Sato and Kanai, 2011). Enzymes from thermophiles can also be used at ambient temperatures and “thermozymes” will often have a much longer half-life at these temperatures than their mesophilic counterparts, giving them a higher total turnover number (TTN) and making them very attractive for industrial use (Rogers and Bommarius, 2011).

Thermozymes can be over expressed in a mesophilic host such as *Escherichia coli*. This gives advantages such as allowing easy purification by differential thermal stability with respect to most of the proteins from the host, as well as a good understanding of the genetic tools and fermentation technology for these organisms (Demirtas, Kolhatkar and Kilbane, 2003).

On the other hand, production in a mesophilic host has its disadvantages. Thermophilic proteins can often be insoluble in mesophilic bacterial cells, forming inclusion bodies (Peng *et al.*, 2012). Sometimes thermozymes that form oligomers, require post-translational modification, proteolytic cleavage or suitable chaperones, are difficult to express in mesophilic hosts (Cava, Hidalgo and Berenguer, 2009). However, most notably there is no opportunity for thermophilic whole-cell biocatalysis while using a mesophilic host, an area which is widely anticipated yet has remained largely undeveloped (Taylor *et al.*, 2011).

### **1.2.2 Thermophile classification, identification and their place in the tree of life**

Thermophiles are organisms which grow at high temperatures, classified as 50 °C and above. The theoretical limit for life is 150 °C (Van Dover, 2000) although to date life has only been found at a maximum temperature of 121 °C (Kashefi and Lovley, 2003). Thermophiles are further categorised into moderate thermophiles, which grow between 40 and 60 °C, extreme thermophiles between 60 and 85 °C, and hyperthermophiles for organisms that can live at greater than 85 °C (Madigan *et al.*, 2012).

Thermophiles have been identified from a variety of natural and man-made sources including geothermal and volcanic areas, deep-sea hydrothermal vents, geothermally heated oil and petroleum reserves, solar-heated soils or sediments, acid mine drainage and acidic effluents, biological wastes and waste treatment plants and self-heated compost piles (Mehta and Satyanarayana, 2013).

Many hyperthermophiles are located on short branches near the base of rooted phylogenetic trees based on 16S rRNA sequences (Glansdorff and Xu, 2004). This observation has led to the hypothesis that the last common ancestor of all extant organisms was possibly a hyperthermophile, and that life originated at high temperatures. However, there is still much debate around this (Stetter, 1996; Farmer, 1998).

For a microbe to grow at high temperature, major components such as proteins, nucleic acids and lipids must be able to resist heat. Thermophiles have various adaptations allowing them to survive and thrive at high temperatures (Mehta and Satyanarayana, 2013)

### **1.2.3 DNA stabilisation in thermophiles**

Several techniques are employed by thermophiles in order to stabilise their DNA at high temperatures. Many hyperthermophiles possess a reverse gyrase, which is a type I DNA topoisomerase that catalyses the positive super coiling of closed circular DNA. Positive supercoiling of DNA is thought to be an important factor in stabilising DNA at high temperatures (Heine and Chandra, 2009).

Other proteins have also been shown to be active in maintaining DNA in hyperthermophiles at high temperature. For example some Euryarchaeotes contain histone-like proteins, which act in a similar way to the core histones of Eukaryotes, winding and compacting DNA into nucleosome like structures allowing DNA to be maintained in a double stranded form at high temperatures (Sandman, Pereira and Reeve, 1998). Other small DNA-binding proteins can bind to the minor groove of DNA non-specifically, increasing the melting temperature of DNA by up to 40 °C (Robinson *et al.*, 1998).

Compatible solutes such as polyamines can also stabilise DNA, in addition to other molecules (Kurz, 2008). Monovalent and divalent salts are able to enhance the stability of DNA as they can screen the negative charge of the phosphate backbone. Salts such as KCl and MgCl<sub>2</sub> have also been shown to protect DNA from depurination and hydrolysis (Marguet and Forterre, 1998).

Some thermophilic bacteria show a higher than average GC content (Marashi and Ghalanbor, 2004). G-C pairs are to be more thermostable due to an additional hydrogen bond, and it has been considered that thermophiles are using a higher GC content to stabilise their DNA (Galtier, Tourasse and Gouy, 1999). However there are also many examples of mesophiles with a high GC content (Dillon *et al.*, 2015), and thermophiles with a more average GC content (Rozanov, Logacheva and Peltek, 2014). An analysis of the GC content of a number of different prokaryotes did not detect a significant correlation of GC content with optimum growth temperature, although a correlation of GC content in structural RNAs with optimum growth temperature was detected (Hurst and Merchant, 2001).

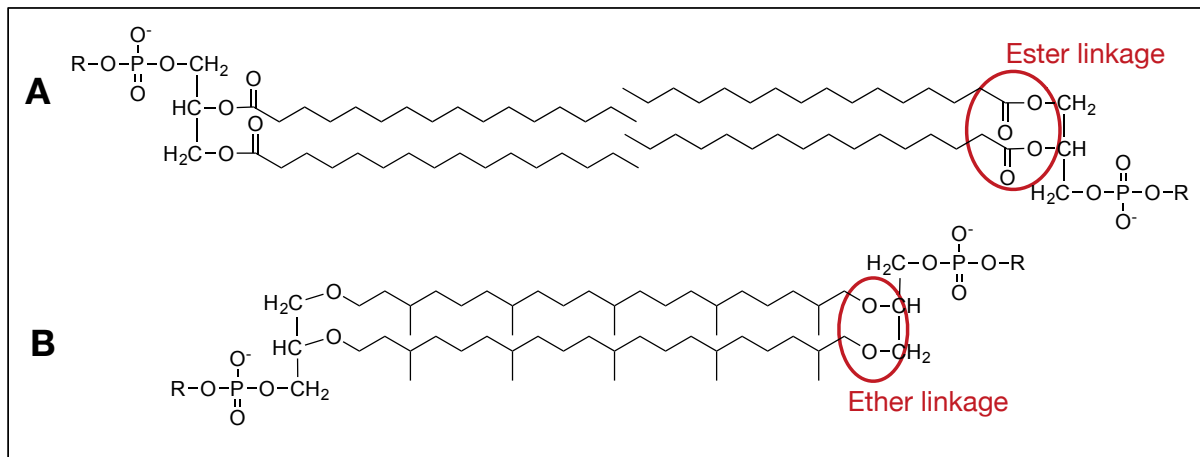
The GC content of an organism has a strong effect on the amino acid composition of their proteins. High GC content may be a result of a preference for certain amino acid residues, which may be involved in coping with elevated temperatures (Wu *et al.*, 2012). For example *Thermus thermophilus*, which has a GC content of 70 %, uses many proline residues to help stabilize its proteins, and these are coded by CCT, CCC, CCA, CCG (Imada *et al.*, 1991; Littlechild *et al.*, 2013).

#### **1.2.4 Lipid bilayer stabilization in thermophiles**

Membrane lipid composition is a major factor in determining growth temperature, and because of this the maximum span of temperatures possible for an organism to grow at is 40 °C. At lower temperatures lipids become more rigid, while at higher temperatures they become more disordered and crystalline. Thermophilic bacteria and archaea have different methods of altering their membrane lipid composition to meet their growth temperature (Mehta and Satyanarayana, 2013).

The lipids of thermophilic bacteria are all ester linked, with the fluidity of the membrane due to the fatty acid chain composition. As temperature is increased, chain length increases, saturation increases and there is an increase in branching (Koga, 2012).

The lipids of archaea are ether linked. These are either diethers forming a bilayer membrane, or tetraethers forming a monolayer membrane. The hydrocarbon chain lengths are always fixed at either 20 or 40 carbons. All hyperthermophiles have a tetraether membrane, and at very high temperatures cyclisation of the acyl chains can be introduced. The failure of bacteria to grow above 80 °C is thought to be due to a lack of tetra ether lipids (Koga, 2012).



**Figure 1-6 – Examples of the lipids of bacteria and archaea.**

- A. The lipids of bacteria have ester linkages. A phospholipid is shown as an example. They form bilayers with the fluidity of the membrane determined by the fatty acid composition.
  - B. The lipids of archaea have ether linkages. They form either diethers forming a bilayer, or tetraethers, as shown in the figure, forming a monolayer. Tetraether lipids allow the membranes of hyperthermophiles to be stable at higher temperatures (Koga, 2012).
- R is used to represent functional groups that might be incorporated as part of the head group.

### 1.2.5 Stabilisation of thermophilic proteins

When proteins are denatured by insults such as elevated temperature, the combination of non-covalent forces holding the protein in its tertiary structure, which includes hydrogen bonds, ion pairs, hydrophobic bonds and Van der Waals interactions, are disrupted. The protein unfolds and is prone to aggregation as exposed hydrophobic residues interact with the hydrophobic residues of other proteins. Chemical modifications such as cysteine oxidation, deamination of asparagine and glutamine residues and peptide bond hydrolysis also can occur (Littlechild *et al.*, 2013).

The proteins of thermophilic organisms must also be adapted to function at high temperature. Generally, these adaptations result in much more stable proteins overall, giving resistance against other factors such as solvents or extremes of pH (Gomes and Steiner, 2004). This makes the enzymes of thermophiles (sometimes called thermozyms) very attractive for industrial processes (Littlechild, 2015). These



adaptions are often unpredictable. However some trends can be observed and are summarised in Table 1-5 (Li, Zhou and Lu, 2005).

Some amino acids are more prone to chemical modification while others have advantageous properties conferring thermostability. Comparisons between proteins from thermophiles and mesophiles show some trends for increases or decreases in certain amino acids, although these adaptions vary between proteins which is not surprising given their different environments and tertiary structures. An increasing frequency of charged residues in thermophilic proteins has been observed, along with a decrease in polar, uncharged residues. This likely increases the number of ionic interactions available to stabilise the structure (Littlechild *et al.*, 2013).

There is a trend for the replacement of glycine with alanine possibly stabilising  $\alpha$ -helices, and lysine with arginine which is less susceptible to chemical attack at high temperatures (Vieille, Zeikus and Vieille, 2001). Cysteine residues are also less common in thermophilic proteins, except where they are required for activity, as they are susceptible to oxidation at high temperatures. Increased frequencies of aromatic and proline residues can offer greater rigidity, conferring thermostability (Fleming and Littlechild, 1997; Korkhin *et al.*, 1999)

Hyper-thermophilic proteins have been shown to have substantially more ion pairs than mesophilic proteins (Aguilar *et al.*, 1997). The ion pairs in thermophiles are often observed clustered together, forming extended ionic networks offering even greater stability than the action of the ionic bonds alone (Isupov *et al.*, 1999).

Proteins fold so as to bury their hydrophobic residues within their core, and this is known to play a major role in protein stability. An average increase in stability of 1.3 kcat mol<sup>-1</sup> has been calculated for each additional methyl group buried in protein folding (Pace, 1992).

Hydrophobic interactions are thought to be key to the thermostability of a number of thermophilic proteins, and increased hydrophobic interactions are observed at subunit interfaces in the crystal structures of many hyperthermophilic proteins. Furthermore, thermophilic proteins are often found to adopt a higher oligomeric state than their mesophilic counterparts (Singleton, Isupov and Littlechild, 1999; Littlechild *et al.*, 2013).

Often a reduction in the frequency and length of external loops is observed in thermophilic proteins (Vieille, Zeikus and Vieille, 2001). In addition to this loops can be better anchored to the rest of the structure through interactions such as ion pairs, hydrogen bonding or hydrophobic interactions. Often *N*- or *C*- termini can be seen anchored to the core or even interacting with each other increasing stability (Auerbach *et al.*, 1997; Littlechild *et al.*, 2013).

**Table 1-5 - Adaptions that can increase thermostability** (Littlechild *et al.*, 2013).

- Increased hydrogen bonding.
- An increase in ionic interactions. Networks of ionic interactions can give greater stability than the sum of the ionic interactions alone.
- Increased hydrophobicity in the core.
- Stabilisation of  $\alpha$ -helices.
- Increased packing to fill internal cavities such as C-terminal extensions and additional secondary structure.
- Binding of metal ions.
- Shorter surface loops, or loops stabilised by interactions with metal ions.
- More rigidity, less flexibility. For example by increased proline content.
- Increased usage of disulphide bonds.

## 1.3 Whole-cell biocatalysis at elevated temperatures

### 1.3.1 The potential for whole-cell biocatalysis at elevated temperatures.

While the enzymes of thermophiles are often employed for industrial biocatalysis, whole-cell biocatalysis in a thermophilic host is an area which is widely anticipated, yet fairly undeveloped (Taylor *et al.*, 2011). Carrying out processes at high temperatures have a number of advantages, yet there are also a number of challenges and these are summarised in Table 1-6 (Frock and Kelly, 2012). Thermophiles are thought to be especially well suited for the consolidated

bioprocessing of lignocellulose and the production of fuels such as bioethanol (Taylor *et al.*, 2011).

<b>Table 1-6 - Advantages and disadvantages of whole-cell biocatalysis at elevated temperatures.</b> (Frock and Kelly, 2012).	
<b>Advantages</b> <ul style="list-style-type: none"> <li>• Reduced risk of contamination.</li> <li>• Recovery of volatile products.</li> <li>• Lower risk of release of viable genetically modified organisms.</li> <li>• Higher mass transfer rates.</li> <li>• Improved solubility of some substrates.</li> <li>• High temperatures more consistent with some chemical processes, allowing integration.</li> <li>• Do not need to cool the reactor.</li> <li>• Some evidence that thermozymes produced homogeneously can work better as biocatalysts (Mei <i>et al.</i>, 2012).</li> </ul>	<b>Disadvantages</b> <ul style="list-style-type: none"> <li>• Energy burden of heating reactor.</li> <li>• Currently lower yields of biomass.</li> <li>• Genetic stability of hosts unknown.</li> <li>• Lower gas solubilities.</li> <li>• Substrate or product lability at elevated temperatures.</li> <li>• Less known about microbial physiology.</li> <li>• Genetic tools are undeveloped.</li> </ul>

Many of the disadvantages in Table 1-6 relate to the undeveloped nature of whole-cell biocatalysis at elevated temperatures and more work is needed in order to overcome these. The major challenge is the lack of genetic tools and these must be developed in order for a thermophile to be considered for whole-cell biocatalysis (Taylor *et al.*, 2011).

### 1.3.2 Potential thermophilic host organisms as cell factories for whole-cell biocatalysis

Several thermophiles have been identified as model organisms at elevated temperatures and these are excellent starting points for developing a host chassis for thermophilic whole-cell biocatalysis. Methodologies for gene inactivation and genetic complementation have been established for several hyperthermophilic archaea, including the euryarchaeon *Pyrococcus furiosus* and a number crenarchaeotic *Sulfolobus* species (Taylor *et al.*, 2011). The thermophilic bacterium *Thermus thermophilus* and a range of *Geobacillus* species are also very promising options with genetic systems available (Cava, Hidalgo and Berenguer, 2009; Kananavičiute and Čitavičius, 2015).

*P. furiosus*, a hyperthermophilic anaerobic archaeon with an optimum growth temperature of 98 °C, is able to take up DNA efficiently by electroporation (Farkas *et al.*, 2012). DNA can be integrated into the chromosome utilising the presence or absence of *pyrF* as a marker (Farkas *et al.*, 2012). An interesting use of *P. furiosus* has been the use of a temperature regulated promoter *PcipA*, which allows the use of resting cells at 72 °C for the expression of heterologous proteins (Basen, Sun and Adams, 2012). *P. furiosus* has also been shown to be useful for carboxylic acid reduction utilising the native aldehyde ferredoxin oxidoreductases and a ferredoxin pool pushed towards the reduced form through the use of hydrogen and hydrogenase enzymes (Hollmann *et al.*, 2012).

*Sulfolobus* species grow in volcanic springs with an optimal growth temperature of 75 – 80 °C at pH 2.0 to 3.0. They can be transformed by electroporation and a number of genetic tools have been developed for various species. The best studied *Sulfolobus* species are *S. islandicus*, *S. acidocaldarius* and *S. solfataricus* (Berkner and Lipps, 2008). The deletion of the *lacS* or *pyrF* genes has allowed genome insertions or deletions to be carried out in these organisms, and would facilitate their use of a host organism for whole-cell biocatalysis (Atomi, Imanaka and Fukui, 2012; Wagner *et al.*, 2012).

*Geobacilli* are Gram-positive, aerobic or facultative anaerobic thermophilic bacteria with an optimum growth temperature between 50 and 60 °C. However, growth temperatures can range from 37 to 80 °C. Similar to *Bacillus* species they are able

to utilise various carbon sources making them an attractive host for biotechnological applications. Electroporation is a common transformation technique for *Geobacilli* although other methods are possible. Plasmid vectors are available for *Geobacilli* as well as thermostable antibiotic resistance cassettes, most commonly kanamycin, which have been used for selection (Kananavičiute and Čitavičius, 2015).

*Thermus thermophilus* is a particularly attractive choice as it is naturally competent, aerobic and able to grow quickly on complex media such as yeast extract and trypticase (Cava, Hidalgo and Berenguer, 2009). *Thermus* is one of the most widely distributed genera of thermophilic bacteria and many of its isolates carry genes for enzymes of great biotechnological potential (Moreno *et al.*, 2005). Strains HB27 and HB8 of *T. thermophilus* have been widely used as laboratory models due to efficient and natural competence and the ability to grow at good rates under laboratory conditions (Koyama *et al.*, 1986). This has allowed the development of a number of genetic tools that are able to function at 60-70 °C, including selectable genes, cloning and expression plasmids, reporter genes and fluorescent localisation markers (Cava, Hidalgo and Berenguer, 2009). *T. thermophilus* has been investigated as a host organism for whole-cell biocatalysis in this study.

## **1.4 *In vitro* cascade reactions and the application of mathematical modelling**

### **1.4.1 The use of isolated enzymes for biocatalysis in vitro.**

The use of isolated enzymes for cell free biocatalysis have a number of advantages as discussed in Section 1.1.9. Pathways of multiple enzymes for the construction of artificial metabolisms *in vitro* is an exciting concept, often called Systems Biocatalysis. Systems Biocatalysis offers the freedom to design the most appropriate sequence of reactions to the target product, free from competing metabolic pathways, kinetic restrictions of physical barriers and regulatory circuits, toxicity problems (Fessner, 2015). *In vitro* biosynthetic systems could surpass the constraints of whole cells or cell lysates, with the potential to become a disruptive bio-manufacturing platform (Zhang, 2015). Well defined enzyme modules might be used to construct new reactions using synthetic biology approaches, with models of

each module allowing quick evaluation of new processes *in silico* (Ladkau, Schmid and Bühler, 2014).

For the use of isolated enzymes to compete with whole-cell biocatalysis enzyme stability is key. The total turnover number (TTN) of an enzyme can be calculated as its turnover number ( $k_{cat}$ ) divided by its half-life at the process temperature, giving an indication of the best enzyme for a process (Rogers and Bommarius, 2011). Higher TTN's allow much more economically efficient processes to be constructed, with the cost of enzyme production relative to the product yield decreasing (Zhang, 2015).

The use of enzymes from thermophiles is an excellent starting point in greater enzyme stability, and engineering and immobilization techniques may be used to improve this (Hickey *et al.*, 2007; Woodley, 2013). As the stability of enzymes is improved and the cost per kg of enzyme production is decreased, the use of enzymes not only for high value pharmaceuticals or specialty chemicals, but also for the production bulk chemicals from renewable feedstocks using isolated enzymes becomes feasible (Zhang, 2015).

However, enzyme production cost is not the only factor in considering a process using isolated enzymes. Oxidoreductases which catalyse many industrially relevant reactions typically require cofactors such as NAD(P)H. For these reactions to be economically competitive a method for cofactor regeneration is essential.

Regeneration can also help drive a reaction to completion and allow the removal of inhibitory cofactor by-products. For cofactor regeneration to be economical and practical the total turnover number (TTN) of the cofactor must also be high. TTNs of  $10^3$  to  $10^5$  are often enough to make a process economically viable, although this depends on the cost of the cofactor and the value of the product. Chemical and electrochemical strategies generally lack the high selectivity needed to reach high TTN's, and for this reason enzymatic reactions are usually the method of choice for cofactor regeneration (Woodyer, Johannes and Zhao, 2003).

Generally NAD(H) functions in reactions involved in biodegradation while NADP(H) is commonly involved in biosynthesis. However, both are involved in a wide range of synthetically useful redox reactions (Woodyer, Johannes and Zhao, 2003). ATP regeneration is also possible but this is an area which is much less developed (Zhao and Van Der Donk, 2003).

### 1.4.2 Cofactor regeneration

Enzymes such as formate dehydrogenase (Berríos-Rivera, Bennett and San, 2002), glucose dehydrogenase (Wong and Whitesides, 1980) and isopropanol alcohol dehydrogenase have successfully been employed for the regeneration of NADH. NADPH-dependent systems are less developed but approaches using glucose 6-phosphate dehydrogenase or an engineered formate dehydrogenase have been used (Kragl *et al.*, 1996; Tishkov *et al.*, 1999; Van Der Donk and Zhao, 2003; Woodyer, Zhao and Van Der Donk, 2005). One of the most promising systems for both NADH and NADPH regeneration is the use of phosphite dehydrogenase (PTDH) from *Pseudomonas stutzeri* utilising phosphite as a sacrificial substrate. This enzyme has been engineered to accept both  $\text{NAD}^+$  and  $\text{NADP}^+$  as substrates as well as for improved thermostability (Woodyer, Van der Donk and Zhao, 2003; McLachlan, Johannes and Zhao, 2007).

Another approach for cofactor regeneration is to exploit the reversible nature of the reactions catalysed by some oxidoreductases for the regeneration of the cofactor. A sacrificial substrate for a reaction in the reverse direction is used to regenerate the cofactor for the reaction of interest (Robins and Osorio-Lozada, 2012). For example, the reduction of ketones to optically active alcohols by ADHs or KREDs can be driven by the use of a sacrificial substrate such as 2-propanol being oxidised in the opposite direction (Schrewe *et al.*, 2013).

Techniques for the regeneration of ATP are less well developed (Zhao and Van Der Donk, 2003), but four enzymes have been employed for the regeneration of ATP from ADP. Phosphoenolpyruvate coupled with pyruvate kinase (Zhao and Van Der Donk, 2003), acetylphosphate coupled with acetate kinase (Yan *et al.*, 2014), creatine phosphate coupled with creatine kinase (Wei and Goux, 1992) and polyphosphate coupled with polyphosphate kinase (PPK) (Restiawaty *et al.*, 2011). Of these, the sacrificial substrate polyphosphate is by far the cheapest while being highly stable, although  $\text{Mg}^{2+}$  chelation remains a challenge (Kameda *et al.*, 2001). For regeneration from AMP, an adenylate kinase or polyphosphate-AMP phosphotransferase may also be used (Resnick and Zehnder, 2000; Woodyer, Johannes and Zhao, 2003).

### 1.4.3 Using isolated enzymes

Isolated enzymes can be used as a crude lysate or can be purified further. Isolated enzymes can be used with no further preparation in a simple batch reaction, however, operational stability and the ability to re-use enzymes can be greatly improved through immobilisation techniques. Enzymes can be bound to a supports or carriers such as synthetic resins, biopolymers or inorganic solids such as silicas (Cau, 2006). Enzymes or carriers can be trapped inside membrane devices such a hollow fibre reactors to facilitate their reuse (Kragl *et al.*, 1996). Alternatively enzymes can be entrapped in a polymer network such as polyacrylamide or silica sol-gel (Sheldon and Van Pelt, 2013). Increasingly there is interest in the use of carrier-free immobilised enzymes using cross-linked enzyme crystals or aggregates (CLECs or CLEAs) (Roy and Abraham, 2004; Sheldon, 2011). This allows for more highly concentrated enzyme activity with high stability and low production costs, making this an attractive option. However every enzyme is different and there no “one size fits all” solution to enzyme immobilisation. Often enzyme immobilisation can depend on the type of reactor to be used such as stirred tank, fixed bed, fluidised bed reactors, different types of membrane reactors or flow reactors (Sheldon and Van Pelt, 2013).

### 1.4.4 The use of process modelling in reaction engineering

Reaction engineering using process models of chemical reactions has for a long time been an efficient and effective method for the design of reaction systems (Charpentier, 2002). In moving from the lab to the process plant a process model can be invaluable allowing decisions to be made on a firm foundation. Often biocatalysis processes do not make this transition well. One of several reasons is the complexity of enzyme kinetics and the lack of available models (Ringborg and Woodley, 2016).

Equations such as those above may be used to construct mechanistic models of a biocatalysis process. The process of building a model can facilitate a fuller understanding the dynamics of a reaction, and uncover phenomena that might otherwise be missed. It can be used as a summary of the available process information (Tufvesson *et al.*, 2013). Furthermore, a validated model can be exploited to optimise a reaction or explore other process options (Rios-Solis *et al.*,



2015). The construction of a model early in the development process can be an invaluable asset in determining the feasibility of a reaction in terms of cost and benefit, or to highlight specifically where enzyme engineering could give substantial gains (Sin, Woodley and Gernaey, 2009). Changes to the process can be simulated *in silico* allowing processes decisions to be well informed (Ringborg and Woodley, 2016).

An alternative to building a mechanistic model for the optimisation of a process is data driven modelling (Solomatine and Ostfeld, 2008). A good example of this is the use of design of experiments methodologies allowing the generation of a response surface. This is a powerful tool and one which is gaining popularity within the synthetic biology community in general (Kennedy and Krouse, 1999). However, it operates as a black box, offering little information on the intricacies of the reaction itself which may offer substantial insight or improvements to a process (Vasić-Rački, Findrik and Vrsalović Presečki, 2011).

Good modelling practice should be used for the construction of process models. Uncertainty and sensitivity analysis can be used to show the confidence in the model's predictions, and to understand which parameters are the most sensitive to error, giving greater uncertainty in the final model prediction (Sin, Gernaey and Lantz, 2009). Where standard errors or 95 % confidence intervals have been reported for a parameter these may be used in the uncertainty analysis. In other cases parameters can be placed into categories of low, medium or high uncertainty with bounds of  $\pm 5\%$ ,  $\pm 25\%$  and  $\pm 50\%$  of the parameter value used respectively (Sin, Woodley and Gernaey, 2009).

#### **1.4.5 Genetic algorithms for optimisation**

Genetic algorithms are a technique for the optimisation of a system towards an object or objectives. The concept was developed in the 1960s and 1970s by Holland and colleagues (Holland, 1975). The idea is based on the theory of evolution, in which weak and unfit species are faced with extinction by natural selection while strong species have more opportunity to pass on their genes to future generations by reproduction (Konak, Coit and Smith, 2006). Genetic algorithms are particularly well suited to multi-objective problems and are good at avoiding local minima or maxima (Deb *et al.*, 2000).

In a genetic algorithm a population of solutions, known as individuals, is ranked against an objective function, with the best performing individuals allowed to survive or reproduce. New individuals are generated by the crossover of successful individuals and by the introduction of mutations. Multiple rounds of selection and reproduction can be carried out to reach an optimised solution (Konak, Coit and Smith, 2006).

Genetic algorithms represent a useful option for the optimisation of process models towards a particular objective function. Examples include achieving a certain level of productivity while minimising cost, minimising the formation of an intermediate or by-product, or achieving a certain yield of product in a specified time frame with minimised cost (Kuhn *et al.*, 2010).

## **1.5 Carboxylic acid reductase enzymes and their use in the construction of enzymatic cascade reactions**

### **1.5.1 Carboxylic acid reductases.**

The reduction of carboxylic acids is challenging. Thermodynamically, COOH is in an energetically favourable state and so shows little reactivity, needing a high level of energy for activation to participate in chemical reactions (Napora-Wijata, Strohmeier and Winkler, 2014). Chemical reductions of carboxylic acids mostly proceed by the formation of an activated derivative via a reaction with complex metal hydrides such as lithium aluminium hydride and sodium borohydride used in stoichiometric amounts, in the absence of water (Nystrom and Brown, 1947). Since aldehydes are more reactive than carboxylic acids, generally the reduction of carboxylic acids proceeds through to the respective alcohol (Bhaskar Khant and Periasamy, 1991).

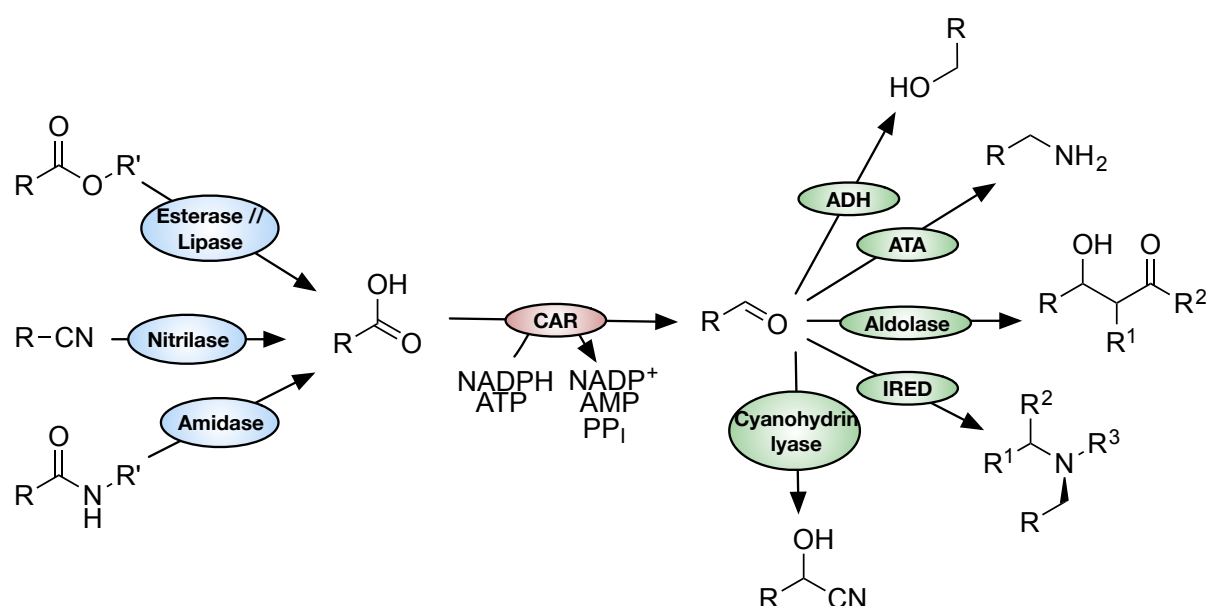
In contrast carboxylic acid reductase (CAR) enzymes offer a green and regio-selective alternative without the use of harsh reducing agents in stoichiometric amounts (Akhtar, Turner and Jones, 2013). These are large, multi-domain enzymes of approximately 130 kDa in size. They feature an N-terminal adenylation domain, a C-terminal thioester reductase domain that likely adopts a Rossmann fold, and a central phosphopantetheine binding domain (Marchler-Bauer *et al.*, 2015). A phosphopantetheine arm must be covalently attached to a conserved serine in this

central domain through the action of a phosphopantetheine transferase for the production of an active enzyme (Venkitasubramanian, Daniels and Rosazza, 2007).

In our study (Chapter 4) on the characterisation of five CAR enzymes, we show that the first step in the proposed reaction mechanism, attack on the alpha phosphate of ATP by a negatively charged oxygen of a carboxylic acid substrate, is likely rate determining. The substrate specificity is broad but similar for each of the CARs tested with electron rich acids favoured substrates. We define operational windows in terms of pH and temperature for each CAR enzyme. We also show that each of the products of the CAR reaction ( $\text{PP}_i$ ,  $\text{NADP}^+$  and AMP) act as inhibitors and that inhibition by  $\text{PP}_i$  gives insight into the ordered binding of substrates at the adenylation domain.

### 1.5.2 Carboxylic acid reductases (CARs) are well suited for the construction of multi-step cascade reactions.

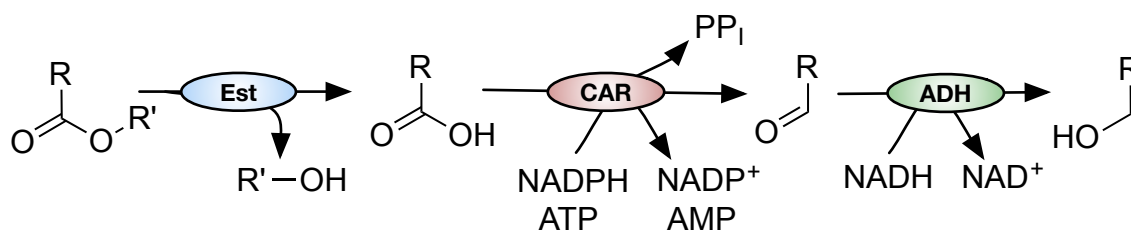
CARs connect many of the enzymes in the enzyme toolbox (Table 1-3), and so allow the construction of attractive multi-step one-pot reactions. Many hydrolase enzymes produce carboxylic acids as one of their products allowing them to be used effectively ahead of CARs, with a range of enzymes accepting aldehydes as one of their substrates (Turner and O'Reilly, 2013). This allows CARs to be used in the construction of a range of multi-step pathways allowing a large combination of reactions to be constructed (France *et al.*, 2016), as shown in Figure 1-7.



**Figure 1-7 - CARs connect many enzyme reactions allowing a host of multi-step reactions to be constructed.**

ADH: alcohol dehydrogenase, ATA: amino transferase, IRED: imine reductase, CAR: carboxylic acid reductase. Cofactors and additional substrates are not shown for clarity.

CARs are well suited for multi-step reactions featuring hydrolase enzymes such as esterases. Nevertheless, chemically the hydrolysis of esters is relatively straightforward, typically catalysed using dilute alkali or acid with heating under reflux (Girard, 1980). However a one-pot chemical multistep pathway for the hydrolysis of an ester followed by the reduction of the resulting carboxylic acid as is unfeasible, as the addition of lithium aluminium hydride does not allow the presence of water (Nystrom and Brown, 1947). In contrast, CARs may be combined with an esterase allowing a one-pot reaction. The addition of an alcohol dehydrogenase to this reaction after the CAR step creates a reaction which demonstrates well the potential of CARs for multi-step biocatalysis, as detailed in Figure 1-8.



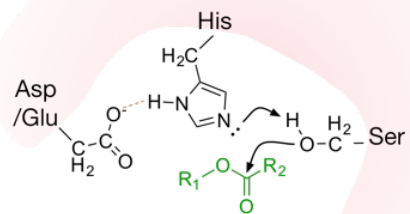
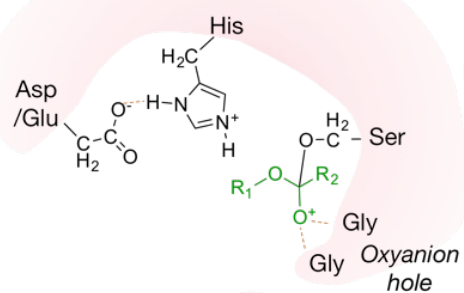
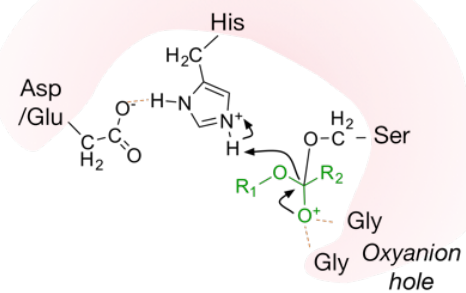
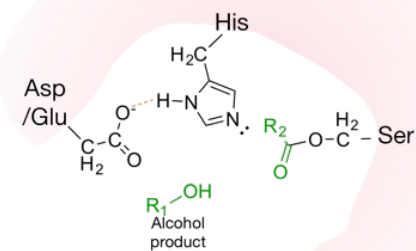
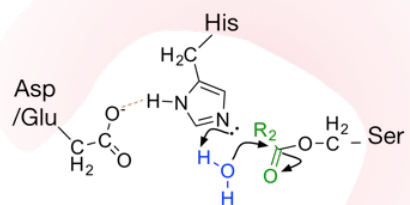
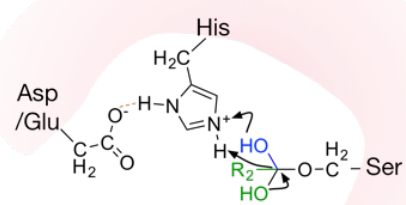
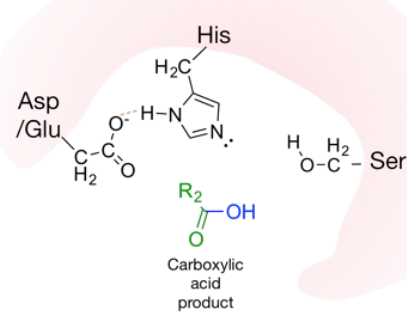
**Figure 1-8 - Reaction scheme for an enzymatic cascade reaction for the hydrolysis and reduction of an ester through to an alcohol**

A carboxylic acid reductase (CAR) may be combined with an esterase (Est) and an alcohol dehydrogenase (ADH) for an attractive one-pot process for the hydrolysis and subsequent reduction of an ester through to its derivative alcohol. R and R' represent any functional group, although substrate specificity is limited by the enzymes used.

### 1.5.3 Esterases

Esterases (EC 3.1.1.1) and their cousins the lipases (EC 3.1.1.3) have received generous attention in the past as they are widespread throughout nature and are very useful for chemical synthesis. Both esterases and lipases catalyse the cleavage of an ester into a carboxylic acid and an alcohol. Esterases can hydrolyse water soluble esters with short acyl chains, while lipases have a preference for insoluble esters with an acyl chain of greater than ten carbons in length (Ferrer *et al.*, 2016).

Most esterases and lipases belong to the  $\alpha/\beta$  hydrolase protein fold superfamily, although examples of enzymes with esterase activity that do not have this fold have been found (Upton and Buckley, 1995; Wagner *et al.*, 2002). Lipases are further differentiated from esterases in that they undergo the processes of interfacial activation in which a hydrophobic lid domain moves away from the activity site in the presence of a minimum concentration of a lipid substrate. Esterases do not show this characteristic and obey classical Michaelis-Menten kinetics (Ferrer *et al.*, 2016).

**A.****1.****2.****3.****4.****B.****5.****6.****7.**

**Figure 1-9 - The reaction mechanism of  $\alpha/\beta$  hydrolase enzymes.**

**A. Acylation step.**

1. Histidine and aspartate/glutamate residues deprotonate the catalytic serine promoting nucleophilic attack by the serine oxygen on the carbonyl carbon of the ester.
2. A tetrahedral intermediate is formed, stabilised by an oxyanion hole.
- 3, 4. An alcohol product is liberated by general acid catalysis by the aspartate/glutamate activated histidine, and the formation of an acyl-enzyme intermediate.

**B. Deacylation step.**

5. A water activated by the histidine residue undergoes nucleophilic attack of the carbonyl carbon of the acyl-enzyme intermediate
6. A second tetrahedral intermediate is formed.
7. The carboxylic acid product is liberated.

Catalysis by esterases which have the classic  $\alpha/\beta$  hydrolase fold generally proceeds via a mechanism similar to that of serine proteases, via a catalytic triad of serine, histidine and aspartate/glutamate in which an acyl-enzyme intermediate is formed, and the alcohol product is released (Kraut, 1977). The acylation step (Figure 1-9A) proceeds via nucleophilic attack on the carbonyl carbon by a catalytically active serine, activated by a histidine and an acidic aspartate or glutamate residue, resulting in a tetrahedral intermediate. An oxyanion hole is crucial in stabilising the carbonyl oxygen in this intermediate. An acyl-intermediate is formed with the release of the alcohol group (Berg, Tymoczko and Stryer, 2010).

The second is known as the deacylation step (Figure 1-9B) in which the acyl-enzyme intermediate is hydrolyzed releasing the carboxylic acid product. Generally, the deacylation step is much slower than the first step and is considered to be rate limiting. The deacylation step proceeds via hydrolysis of the acyl-intermediate by a water molecule, again with the formation of a tetrahedral intermediate before the release of the carboxylic acid product (Voet and Voet, 2004).

Esterases have been used for a variety of biotechnology applications (Levisson, van der Oost and Kengen, 2009), including the synthesis of optimally pure compounds for food, agriculture and pharmaceutical industries (Romano *et al.*, 2015). They also have found uses in the degradation of natural materials, pollutants, plastics and other waste (Gomes and Steiner, 2004). Esterases are often shown to possess excellent stability in organic solvents (Klibanov, 2001), and to have a broad substrate specificity with high region- and stereo- selectivity, making them attractive for industrial use (Somers and Kazlauskas, 2004; Yang *et al.*, 2015).

#### **1.5.4 Alcohol dehydrogenases**

Alcohol dehydrogenases (ADHs, (EC 1.1.1.1) and keto-reductases (KREDs, EC 1.1.1.2), catalyse the conversion of ketones or aldehydes to alcohols, or vice versa (Kaluzna, David Rozzell and Kambourakis, 2005). While ADHs and KREDs catalyse similar reactions, they are different classes with different overall folds (Hyndman *et al.*, 2003; Knoll, 2008). This is one of the most common redox reactions in organic chemistry, however synthetic chemical methods require the use of toxic metals and complex hydrides (Brown and Chaikin, 1949). In contrast, biotransformations offer mild reaction conditions with excellent specificity making this an attractive class of



enzyme for use in synthetic chemistry (Goldberg *et al.*, 2007). The majority of the dehydrogenases or reductases that can be used for these biotransformations require nicotinamide cofactors, and so their regeneration must be considered for the use of these enzymes (Kroutil *et al.*, 2004).

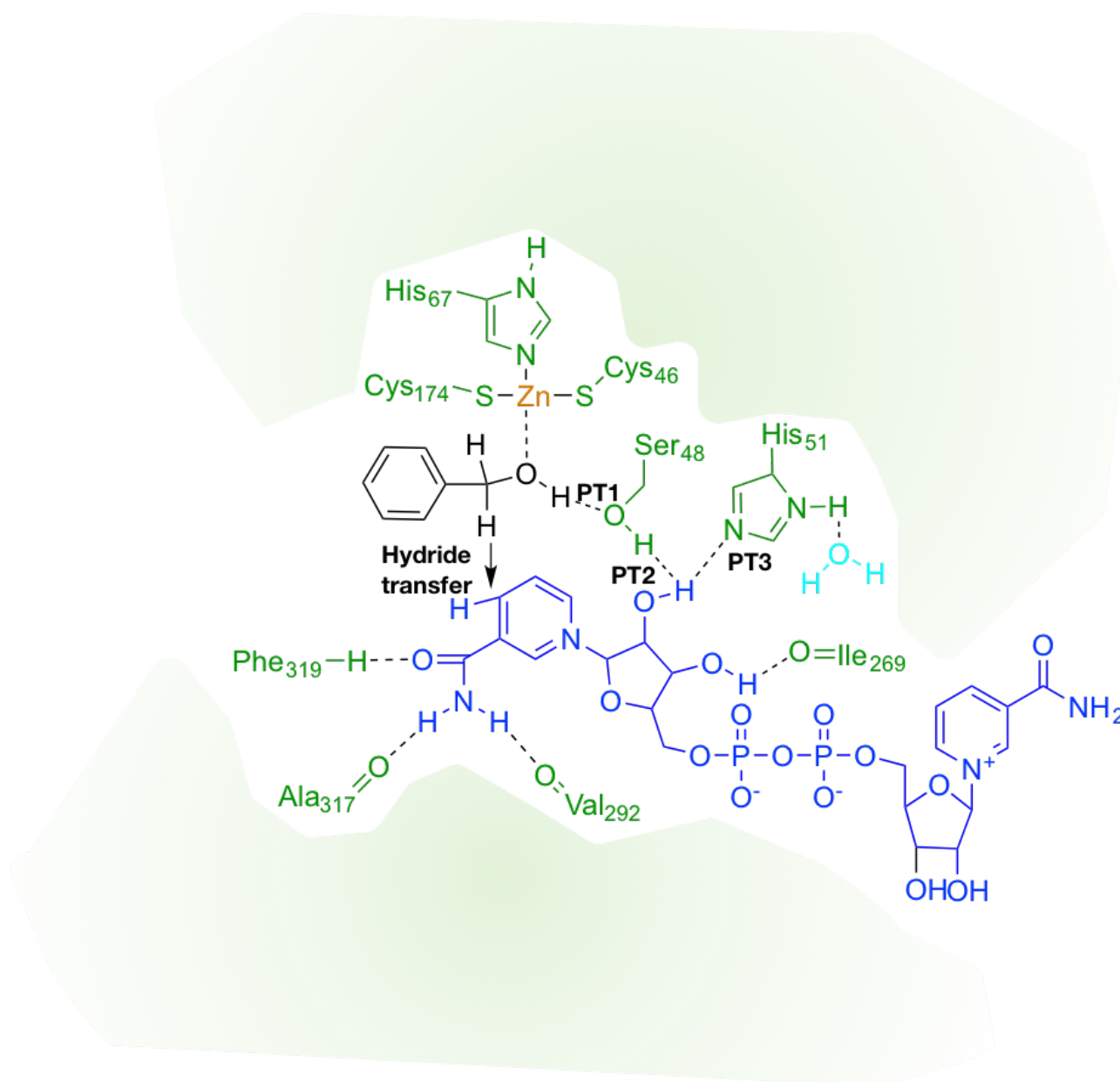
ADHs are distributed widely in nature and have been isolated from species throughout the three domains of life (Guy, Isupov and Littlechild, 2003). The ADH's can be classified according to three classes. Type I ADHs, or medium chain ADHs are dimers or tetramers with each subunit containing approximately 370 amino acid residues (Reid and Fewson, 1994). These ADH's may also be split into non-zinc dependent and zinc dependent, where often both a structural zinc and a catalytic zinc is present (Knoll, 2008). Type II ADH's, or short chain ADH's, are generally around 250 amino acid residues in length and generally do not contain metals. Type III ADH's, or long chain ADH's, are between 380 and 900 amino acids in length and are generally iron dependent. The type III ADH's show little sequence homology to either of the other two classes of ADH (Reid and Fewson, 1994).

Classic examples of ADH's include horse liver ADH (HLADH) and yeast ADH1 (YADH1). Both of these enzymes are medium chain zinc dependent ADH's and are fairly homologous to each other, except for a deletion of 21 amino acid residues from the catalytic domain of yeast ADH1, as well as some other sequence gaps and insertions (Raj, Ramaswamy and Plapp, 2014). Medium chain zinc-dependent ADH's have a conserved tertiary structure made up of two domains, a cofactor binding domain and a catalytic domain. Most are active as dimers or tetramers, with a loop segment in the catalytic domain mediating the formation of the quaternary structure (Knoll, 2008).

The reaction mechanism of HLADH is generally agreed to proceed as follows (Agarwal, Webb and Hammes-Schiffer, 2000), as is shown in Figure 1-10.

1. Binding of  $\text{NAD}^+$
2. Binding of alcohol by coordination with the catalytic zinc
3. Deprotonation of the alcohol via a proton relay in which the proton is transferred via a hydrogen bonded system of hydroxyl groups on Ser48 and NAD before reaching His51. The proton is then transferred to the aqueous solvent. This leads to a zinc bound alkoxide ion
4. Hydride transfer from the alkoxide ion to  $\text{NAD}^+$ , leading to a NADH and a zinc bound aldehyde or ketone
5. Release of the aldehyde
6. Release of NADH

HLADH and YADH1 have been shown to follow a Theorell-Chance mechanism (which is a specific case of an ordered bi-bi reaction), in which there is no significant build up of EAB or EPQ intermediates (Kuby, 1991). This occurs as there is direct hydride transfer between the substrate and  $\text{NAD}^+$  (Agarwal, Webb and Hammes-Schiffer, 2000).



**Figure 1-10 - Mechanism of horse liver alcohol dehydrogenase.**

The mechanism of alcohol oxidation in the active site of horse liver alcohol dehydrogenase. Benzyl alcohol is shown as the substrate in black, with NAD<sup>+</sup> in blue and the key residues of the protein in green. The catalytic zinc, coordinated by two cysteines and a histidine is shown in orange, and a water molecule in light blue. PT1 to 3 refers to the proton transfer relay to His51. Adapted from (Agarwal, Webb and Hammes-Schiffer, 2000).

## 1.6 Aims and objectives

As discussed in section 1.1.12, multistep biocatalysis is an attractive prospect industrially. This work aims to investigate the use of a multistep enzymatic cascade reaction featuring an esterase, a carboxylic acid reductase and an alcohol dehydrogenase for the hydrolysis and subsequent reduction of an ester through to an alcohol.

Enzymes which can take a proof of concept industrially relevant chemical (methyl *p*-toluate) to its derivative alcohol will be identified for the construction of this cascade. In the process new enzymes are to be explored and characterised to determine their potential for biocatalysis. Carboxylic acid reductases are of particular interest industrially and a thorough biochemical characterisation of these enzymes from available genomes will be carried out (Chapter 4). Novel thermophilic esterases will also be investigated, with the possibility of collaborations with other work being carried out at the University of Exeter (Chapter 3). Where possible the use of thermostable enzymes will be explored offering high operational stability and allowing their expression in a thermophilic bacterium.

As discussed in section 1.3, whole-cell biocatalysis at elevated temperatures is an attractive yet undeveloped prospect. A synthetic biology approach will be used in developing the use of the thermophilic bacterium *T. thermophilus* for whole-cell biocatalysis. This will include the design and characterisation of parts such as expression vectors, promoters, ribosome binding sites and optimised protein coding sequences. Utilising this approach enzymes to carry out the cascade reaction will be expressed in *T. thermophilus* to evaluate its use as a host organism for whole-cell biocatalysis (Chapter 6).

Following the characterisation of the enzymes utilised in the construction of the cascade reaction, this knowledge can be used in the design of an *in vitro* reaction. In combination with characterised enzymes for cofactor regeneration, mathematical modelling of the cascade reaction will be used to understand the dynamics of the multi-enzyme system, as discussed in section 1.4. The model can then be used to optimise the process and explore potential reaction configurations (Chapter 5).

In summary, this work seeks to characterise novel enzymes for inclusion into the enzyme toolbox for synthetic chemistry (Chapters 3 and 4), and combine them into a multistep cascade reaction to investigate both the use of isolated enzymes (Chapter 5), and the use of a thermophilic host organism for whole cell biocatalysis (Chapter 6).



## **Chapter 2 - General Materials and Methods**

## 2.1 Materials

### 2.1.1 Bacterial strains

The bacterial strains used in this study are shown in Table 2-1.

Table 2-1 – Bacterial strains used in this study	
Strain	Use
5-alpha <i>E. coli</i> (NEB)	High transformation efficiency cloning strain
BL21(DE3) <i>E. coli</i> (NEB)	Expression strain
Rosetta 2 <i>E. coli</i> (Novagen)	Expression strain derived from BL21(DE3)
RipL <i>E. coli</i> (Agilent Technologies)	Expression strain derived from BL21(DE3)
<i>Thermus thermophilus</i> HB27:: <i>nar</i>	<i>T. thermophilus</i> HB27:: <i>nar</i> , a strain of HB27 with the addition of the <i>nar</i> element transferred from <i>T. thermophilus</i> HB8, was used in all <i>T. thermophilus</i> work. This strain was kindly gifted by José Berenguer, and is required for the use of the P <sub>nar</sub> promoter in an HB27 strain.



### 2.1.3 Media

The components of media not described in other places is shown in Table 2-2 (Sambrook and Russell, 2012).

Table 2-2 - Growth media used in this study.	
Name	Components
LB media (Luria-Bertani Broth)	10 g/L peptone, 5 g/L NaCl, 5 g/L yeast extract
SOC (Super Optimal broth with Catabolite repression)	20 g/L tryptone, 5 g/L yeast Extract, 4.8 g/L MgSO <sub>4</sub> , 3.603 g/L glucose, 0.5 g/L NaCl, 0.186 g/L KCl

### 2.1.4 Antibiotics

Antibiotics were prepared at 1000x concentration in distilled water, except chloramphenicol which was prepared in ethanol. Stocks were stored at -20 °C and fully defrosted before use. Final working concentrations are shown in Table 2-3.

Table 2-3 - Antibiotics used in this study.	
Antibiotic	Working concentration
Kanamycin	50 µg / ml
Kanamycin for use with <i>T. thermophilus</i>	30 µg / ml
Chloramphenicol	50 µg / ml
Ampicillin	100 µg / ml
Streptomycin	50 µg / ml
Spectinomycin	50 µg / ml

### 2.1.5 Chemicals

Unless otherwise specified, all chemicals were purchased from Sigma-Aldrich (Gillingham, UK), and were of the highest purity available.

### 2.1.6 Antibodies

Antibodies used in western blots in this study are shown in Table 2-4. All were used at a 1:1000 dilution.

Table 2-4 – Antibodies used in this study
<b>Primary Antibodies</b>
Chicken Anti-DDDDK (FLAG) tag antibody (Abcam)
Mouse anti-c-Myc Monoclonal Antibody (Life Technologies)
Mouse anti-penta-His Antibody (Qiagen)
<b>Secondary Antibodies</b>
IRDye 680RD Goat anti-Mouse IgG (H+L) (LI- COR Biosciences)
IRDye 800CW Donkey anti-Chicken IgG (H + L) (LI-COR Biosciences)

## 2.2 DNA methods

### 2.2.1 Plasmid preparation and isolation

Plasmids were prepared from overnight cultures of *E. coli* grown in 10 ml LB media with the appropriate antibiotics (Table 2-3). A QIAprep Spin Miniprep Kit (Qiagen)

was used for plasmid purification following the manufacturer's instructions. DNA was eluted in 50  $\mu$ l distilled water.

### 2.2.2 High fidelity PCR

For PCR reactions requiring high fidelity, where the PCR product would be incorporated into a vector, high fidelity PCR using Phusion polymerase (Fermentas) was performed. A general recipe is shown in Table 2-5.

<b>Table 2-5 – PCR recipe for use with Phusion polymerase</b>	
<b>Component</b>	<b>Vol per reaction (<math>\mu</math>l)</b>
5x High GC buffer	4
Distilled water	11.2
dNTPs at 2.5 mM	1.6
Primer Fwd (10 $\mu$ M)	1
Primer Rev (10 $\mu$ M)	1
DMSO	0.6
Template	0.5
Phusion polymerase (1 U/ $\mu$ L)	0.1
Total	20

Where appropriate a master mix was made for multiple PCR reactions. PCR reactions were performed in an Eppendorf Nexus GSX1 or Biorad thermocycler T100 using the following program:

- 98 °C for 1 minute
- 35 cycles of:
  - 98 °C for 30 seconds
  - 55 °C for 30 seconds (other annealing temperatures were used where appropriate)
  - 72 °C for 30 seconds per 1 kb
- 72 °C for 5 minutes

### 2.2.3 PCR for cloning confirmation

For analytical PCR reactions where the resulting product would only be analysed by agarose gel electrophoresis, such as colony PCR, DreamTaq Green PCR Master Mix (Thermo Scientific) was used. A general recipe is shown in Table 2-6.

<b>Table 2-6 – PCR recipe for use with DreamTaq Green PCR Master Mix</b>	
<b>Component</b>	<b>Vol per reaction (µl)</b>
2x DreamTaq Green PCR Master Mix	5
Distilled water	3.6
Primer Fwd (10 µM)	0.5
Primer Rev (10 µM)	0.5
DMSO	0.3
Template (or water for colony PCR)	0.1
Total	10

Where appropriate a master mix was made for multiple PCR reactions. For colony PCR reactions, colonies were added directly to the PCR reaction. PCR reactions were performed in an Eppendorf Nexus GSX1 or Biorad thermocycler T100 using the following program:

- 94 °C for 1 minute (5 minutes for colony PCR)
- 40 cycles of:
  - 94 °C for 30 seconds
  - 55 °C for 30 seconds (other annealing temperatures were used where appropriate)
  - 68 °C for 1 minute per 1 kb
- 68 °C for 5 minutes

#### **2.2.4 Agarose gel electrophoresis**

DNA fragments were analyzed using agarose gel electrophoresis, separating DNA by size. Generally 1.0 % agarose gels were prepared by heating 0.6 g of agarose in 100 ml 1x TAE buffer until dissolved. 1x TAE buffer was prepared from a 50x TAE buffer stock. 50x TAE buffer was made up of: 142 g/L Tris base, 57 ml/L acetic acid and 19 g/L EDTA.

2.5 µl of Midori Green Advance (Nippon Genetics) was added while cooling the agarose before pouring the agarose into a gel tray with a comb, where it was allowed to set. DNA samples were loaded onto the gel with an appropriate DNA ladder. The gel was run in a Sub-Cell GT Cell (Bio-Rad) at 100 V in 1x TAE buffer for 45 minutes before visualization under UV light using a Bio-Rad BioDoc-It imaging system.

#### **2.2.5 Gel extraction or DNA purification**

Where required, DNA fragments from PCR or restriction digests were extracted from agarose gels and purified using a QIAquick Gel Extraction kit (Qiagen) following the manufacturer's instructions. Elution was carried out in 30 µl distilled water. For the purification of DNA from PCR reactions, SureClean (Bioline) was used following manufacturer's instructions, with DNA suspended in the same volume of distilled water.

### **2.2.6 Restriction digests**

Restriction digests were carried out using up to 1000 ng of purified DNA in a 20  $\mu$ l reaction volume. 1  $\mu$ l each of FastDigest enzymes (Fermentas) were used, and where required 1  $\mu$ l of FastAP Thermosensitive Alkaline Phosphatase (Fermentas) or 1  $\mu$ l *DpnI* FastDigest (Fermentas). Reactions were carried out at 37 °C for 30-60 minutes, followed by 10 minutes at 80 °C. Digests were analysed on agarose gels followed by gel extraction of the desired fragment, or used directly.

### **2.2.7 Ligations**

Restriction digested vector and insert fragments were ligated at a standard molar ratio of 1:3 vector to insert, with between 10-100 ng of vector DNA. Ligation reactions were carried out in a volume of 10  $\mu$ l using T4 DNA Ligase (Thermo Scientific). Reactions contained 1  $\mu$ l 10X ligase buffer, 0.1  $\mu$ l T4 ligase and a total of 8.9  $\mu$ l of distilled water, vector and insert. Ligations were carried out for 20 minutes at room temperature before being placed on ice ready for transformation into a chemically competent cloning strain of *E. coli*.

### **2.2.8 *E. coli* transformation**

Chemically competent strains were prepared by laboratory technicians or purchased ready to use. Competent cells were stored at -80 °C and defrosted on ice. 1-5  $\mu$ l of ligation mixture was then added to 50  $\mu$ l of competent cells and incubated on ice for 30 minutes. Cells were heat-shocked in a water bath at 42 °C for 30 seconds before cooling on ice for 2 minutes. 200  $\mu$ l of LB media or SOC was added and cells were incubated at 37 °C with shaking at 220 RPM for one hour. Cells were plated out onto LB plates containing the appropriate antibiotic, and where pNIC28-BSA4 was used as the cloning vector the addition of 5 % (w/v) sucrose.

## **2.3 Protein purification and analysis**

### **2.3.1 General protocol for protein expression in *E. coli***

Expression plasmids were transformed into the BL21(DE3) strain or other derivative strains of *E. coli*, and glycerol stocks produced. A 10 ml overnight culture was set up

in LB media with the appropriate antibiotics added at 50 µg / ml, and incubated at 30 °C with shaking at 220 rpm. 2 ml of overnight culture was used to inoculate 1 L of LB with the addition of 50 µg / ml of the relevant antibiotics, in a 2 L Erlenmeyer flask, incubated at 37 °C with shaking at 220 rpm. OD<sub>600 nm</sub> readings were taken periodically after the first 3 hours of incubation until OD<sub>600 nm</sub> reached between 0.4 and 0.8. At this point temperature was dropped to 20 or 30 °C with the addition of 100 µM IPTG. Cultures were left overnight for approximately 18 hours before harvesting the cells by centrifugation at 4,700 g for 30 minutes at 4°C, and re-suspended in 25 ml of nickel column wash buffer (25 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole) per 0.5 L of culture. Cell suspensions were either frozen at -20 °C or processed immediately for purification.

### **2.3.2 Cell lysis by sonication**

Cells were lysed by sonication on ice using a Soniprep 150 sonicator (MSE, London, UK), by six cycles of 25 seconds sonication at an amplitude of 10 µm, followed by 35 seconds resting. The insoluble fraction of the cell lysis sample was removed by centrifugation at 20,000 g for 30 minutes at 4°C.

### **2.3.3 Protein purification**

Soluble cell extract was loaded onto 1 ml HisTrap-FF columns (GE Healthcare), pre-equilibrated with nickel column wash buffer, using an ÄKTApurifier system. Columns were washed with 20 ml of wash buffer before elution of bound protein by raising the concentration of imidazole from 20 mM to 250 mM over a 10 ml gradient, with 1.5 ml fractions collected throughout.

Fractions were analysed by SDS-PAGE, and those shown to contain the protein of interest were concentrated to less than 2 ml by ultrafiltration, using a Vivaspin 20 (Sartorius) with a 10 kDa MW cut off. Up to 2 ml was then loaded and run through a HiLoad 16/600 Superdex 200 gel filtration column, pre-equilibrated with GF buffer (25 mM Tris-HCl pH 7.5, 100 mM NaCl), again using an ÄKTApurifier system. Fractions were collected throughout a 120 ml elution in GF buffer, and those thought to contain the protein of interest analysed by SDS-PAGE, before being pooled and concentrated to between 1 mg/ml and 8 mg/ml by ultrafiltration as previously.

### 2.3.4 Protein quantification

Protein concentration was determined by measuring OD<sub>280 nm</sub> using a Nanodrop 2000c (Thermo Scientific), blanked against GF buffer, and utilizing an extinction coefficient and MW calculated by protein sequence analysis by the Expasy ProtParam tool (Gasteiger *et al.*, 2005) (<http://web.expasy.org/protparam/>).

### 2.3.5 SDS-PAGE

SDS-PAGE was used to separate proteins by size for visualization. ExpressPlus PAGE Gels (GenScript) were used with Tris-MOPS (SDS) running buffer consisting of: 6.06 g/L Tris base, 10.46 g/L MOPS, 1 g/L SDS and 0.3 g/L EDTA. Samples were prepared for analysis by dilution in an equal volume of 2X-SDS-PAGE buffer (100 mM Tris-HCl pH 6.8, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (w/v) glycerol, 2% (v/v)  $\beta$ -mercaptoethanol), and heated at 100°C for 10 minutes. 1 – 20  $\mu$ l of each sample was loaded onto the gel. 10  $\mu$ l of a protein ladder such as Spectra Multicolor Broad Range Protein Ladder (Thermo Scientific) was also included. Gels were using a Mini-Protean Electrophoresis System (Bio-Rad), at 150 V for 1 hour. Gels were stained with InstantBlue stain (Expedeon) for 1 hour, and washed with water.

### 2.3.6 Western blot

Western blots were carried out following SDS-PAGE electrophoresis without staining, using a Pierce G2 Fast Blotter (Thermo Scientific) and an iBind Western System (Thermo Scientific). Nitrocellulose (0.22  $\mu$ m pore size, Sartorius Stedim) was cut to size and soaked in 1-Step Transfer Buffer (Thermo Scientific) for 30 minutes with four Western Blotting Filter Papers (7 cm by 8.4 cm, 0.83 mm thickness, Thermo Scientific).

A “sandwich” was constructed of two filter papers, nitrocellulose sheet, SDS-PAGE gel (plastic casing removed) and finally two more filter papers. This was placed onto the Pierce G2 Fast Blotter and the lid attached. Transfer was carried out at 25 V, 1.3 A, for 30 minutes. During the transfer the iBind mix and antibodies were prepared. Following the transfer the nitrocellulose sheet was removed and soaked in 6 ml of iBind mix, ready for the next step.



iBind mix was prepared consisting of 6 ml iBind buffer, 300  $\mu$ l iBind additive and 23.7 ml distilled water. Antibody solutions were prepared at a 1 in 1000 dilution in 2 ml of iBind mix for both primary and secondary antibodies.

An iBind card was placed into the iBind system, and 5 ml of iBind mix pipetted covering the card. 1 ml was pipetted into the center of the card. The nitrocellulose membrane was placed facing down onto the center of the card. The iBind system was clipped shut, and iBind solutions added in slots 1 to 4 as follows:

1. 2 ml primary antibody solution,
2. 2 ml iBind mix,
3. 2 ml secondary antibody solution,
4. 6 ml iBind mix.

The iBind system was left for 2.5 hours before visualization using an Odyssey CLx Imaging System (LI-COR).

## **2.4 Other methods**

### **2.4.1 HPLC method for separation of methyl *p*-toluate, *p*-toluic acid, *p*-tolualdehyde, and *p*-tolyl alcohol**

An Eclipse Plus C18 column with a particle size of 3.5  $\mu$ m, measuring 4.6 x 100 mm, was used. The column was run at 60 °C on the following method using two buffers, buffer A: 95 % H<sub>2</sub>O, 5 % (v/v) acetonitrile, 0.1 % (v/v) trifluoroacetic acid, and buffer B: 5 % H<sub>2</sub>O, 95 % (v/v) acetonitrile, 0.1 % (v/v) trifluoroacetic acid. 3  $\mu$ l of sample was injected and eluted on a gradient from 0 to 100 % buffer B over 10 minutes. Buffer B was maintained at 100 % for a further 2 minutes before the column was re-equilibrated with buffer A for 2 minutes before the next run.



**Chapter 3 - Structural and biochemical  
characterisation of *Archaeoglobus  
fulgidus* esterase reveals a bound CoA  
molecule in the vicinity of the active  
site.**

### 3.1 Authors

Christopher Sayer<sup>1</sup>, William Finnigan<sup>1</sup>, Michail N. Isupov<sup>1</sup>, Mark Levisson<sup>2</sup>, Servé W. M. Kengen<sup>2</sup>, John van der Oost<sup>2</sup>, Nicholas J. Harmer<sup>1</sup>, Jennifer A. Littlechild<sup>1\*</sup>

<sup>1</sup>The Henry Wellcome Building for Biocatalysis, Biosciences, University of Exeter, Stocker Road, Exeter EX4 4QD, UK

<sup>2</sup> Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708WE, Wageningen, The Netherlands

### 3.2 Preface for inclusion in this thesis

This chapter is made up of the reformatted manuscript for a paper published in Scientific Reports in April 2016, volume 6. The esterase in this paper, AF-Est2, makes up the first step in the enzymatic cascade reaction investigated throughout this thesis, shown in Figure 1-1. It was chosen for the cascade due to its ability to accept methyl *p*-toluate as a substrate, and its good thermostability allowing its use in *Thermus thermophilus*. A thorough biochemical characterisation was an important step in developing this cascade reaction, and was combined with the structural work on AF-Est2 carried out by Chris Sayer for publication.

All biochemical work was performed by William Finnigan, except assays for activity against acetyl-CoA and succinyl-CoA which were performed by Chris Sayer. Sections 3.5.1 to 3.5.6 including all relevant figures and tables, and 3.6.1 to 3.6.8 written by William Finnigan, edited by all authors. Other sections were written by Chris Sayer or Michail N. Isupov, edited by all authors.

The manuscript has been formatted to match the formatting of this thesis.

References have been combined with those in the other chapters of this thesis for a final reference chapter. Figure and table numbers have also been edited for inclusion in the thesis.

### 3.3 Abstract

A new carboxyl esterase, AF-Est2, from the hyperthermophilic archaeon *Archaeoglobus fulgidus* has been cloned, over-expressed in *Escherichia coli* and biochemically and structurally characterized. The enzyme has high activity towards short- to medium-chain *p*-nitrophenyl carboxylic esters with optimal activity towards the valerate ester. The AF-Est2 has good solvent and pH stability and is very thermostable, showing no loss of activity after incubation for 30 min at 80 °C. The 1.4 Å resolution crystal structure of AF-Est2 reveals Coenzyme A (CoA) bound in the vicinity of the active site. Despite the presence of CoA bound to the AF-Est2 this enzyme has no CoA thioesterase activity. The pantetheine group of CoA partially obstructs the active site alcohol pocket suggesting that this ligand has a role in regulation of the enzyme activity. A comparison with closely related  $\alpha/\beta$  hydrolase fold enzyme structures shows that the AF-Est2 has unique structural features that allow CoA binding. A comparison of the structure of AF-Est2 with the human carboxyl esterase 1, which has CoA thioesterase activity, reveals that CoA is bound to different parts of the core domain in these two enzymes and approaches the active site from opposite directions.

### 3.4 Introduction

*Archaeoglobus fulgidus* is an anaerobic heterotrophic sulfate reducing archaeon (Klenk *et al.*, 1997) that grows at temperatures between 60 and 95 °C, with optimal growth at 83 °C. The enzymes from thermophilic microorganisms have high potential for use as new industrial biocatalysts due to their naturally high stability to temperature and organic solvents (Littlechild *et al.*, 2007; Littlechild, 2015).

The esterases are a class of enzymes that catalyse the cleavage of ester bonds and have been extensively used in many important biotechnology applications (Levisson, van der Oost and Kengen, 2009). These include the synthesis of optically pure compounds in the agriculture, food and pharmaceutical industries (Romano *et al.*, 2015) and the degradation of natural materials and pollutants, plastics and other industrial waste products (Gomes and Steiner, 2004; Wei, Oeser and Zimmermann, 2014). Features such as high stability in organic solvents (Klibanov, 2001), broad substrate specificity (Yang *et al.*, 2015) and high regio- and stereo-selectivity (Somers and Kazlauskas, 2004) make these enzymes attractive biocatalysts. Specific examples include the use of aryl esterases in the development of flavours in the food and beverage industries (Lomolino *et al.*, 2003), and the use of the *Bacillus* carboxyl esterase NP for the production of the nonsteroidal drug naproxen (Quax and Broekhuizen, 1994).

The lipolytic enzymes are made up from two main groups (Schrag and Cygler, 1997): the true lipases (EC 3.1.1.3; triacylglycerol lipases) and the esterases (EC 3.1.1.1; carboxylesterases). Enzymes of both groups catalyse the cleavage of an ester into a carboxylic acid and an alcohol. The esterases hydrolyse water soluble esters with a short fatty acid chain, while lipases show preference for insoluble fatty acid esters with a chain length greater than ten carbon units. The esterases obey classical Michaelis-Menten kinetics and have a relatively open active site. In contrast, the lipases use the process of interfacial activation where a hydrophobic lid domain moves away from the active site in the presence of a minimum concentration of a lipid substrate. Most lipolytic enzymes belong to the  $\alpha/\beta$  hydrolase protein fold superfamily. However, esterase activity has been reported for enzymes with a  $\beta$  lactamase fold (Wagner *et al.*, 2002) and an  $\alpha/\beta/\alpha$  hydrolase fold (Upton and Buckley, 1995) or as a side activity for the carbonic anhydrase enzymes (Host,

Martensson and Jonsson, 2006). The ESTHER database (Lenfant *et al.*, 2013) divides the  $\alpha/\beta$  hydrolase enzymes into over 140 families and superfamilies which are further assigned to groups C, H, L, and X.

The proposed mechanism of catalysis by the  $\alpha/\beta$  hydrolase lipolytic enzymes resembles the serine protease mechanism (Blow, Birktoft and Hartley, 1969). This involves the substrate binding with the carbonyl oxygen adjacent to the oxyanion hole. The catalytic acidic residues and histidine activate the catalytic serine hydroxyl which performs nucleophilic attack on the carbonyl carbon of the scissile bond to produce the tetrahedral intermediate. Subsequently the alcohol product is released and an acyl-enzyme complex is formed. After an attack by a water molecule another tetrahedral intermediate is formed which resolves to release the carboxyl product and the free enzyme. The catalytic serine residue in the  $\alpha/\beta$  hydrolase fold esterases is usually located in a tight nucleophilic elbow with the consensus sequence Gly-X-Ser-X-Gly, although deviations from this consensus have been reported (Bourne, Isupov and Littlechild, 2000).

Many examples of thermophilic esterases have previously been described, including enzymes from *Thermotoga maritima* (Levisson, van der Oost and Kengen, 2007; Levisson *et al.*, 2012), *Geobacillus stearothermophilus* (Liu *et al.*, 2004, 2007), *Alicyclobacillus acidocaldarius* (Mandrich *et al.*, 2008), *Thermobifida fusca* (Billig *et al.*, 2010), *Sulfolobus tokadaii* (Angkawidjaja *et al.*, 2012), *Rhizomucor miehei* (Yang *et al.*, 2015), amongst others. To date two esterases and a lipase from *A. fulgidus* have been reported. The AFEST (Manco *et al.*, 2000) (locus tag: AF1716) is a member of the hormone-sensitive-lipase family and has a preference for the substrate *p*NP-caprylate. The crystal structure of this enzyme (PDB: 1JJI) revealed several thermophilic adaptations with an increased secondary structure content, reduction in loop extensions and an increase in intramolecular ion pairs compared to a mesophilic homologue (De Simone *et al.*, 2001). The second esterase Est-AF (Kim, Lee and Ryu, 2008) (locus tag: AF2336) showed optimal activity towards *p*NP-butyrate and no activity for substrates with a carboxyl group larger than caprylate. This enzyme has been applied in the resolution of ketoprofen ethyl esters. Extensive mutagenesis of Est-AF resulted in an enzyme with improved activity for the production of the S-enantiomer of ketoprofen (Kim *et al.*, 2015). The biocatalytic synthesis of poly( $\delta$ -valerolactone) using AFEST has also recently been

demonstrated (Cao *et al.*, 2012). The alkaliphilic *A. fulgidus* lipase (Chen *et al.*, 2009) (locus tag: AF1763) is also thermostable at 70-90 °C.

Here, we report the biochemical and structural characterisation of a third esterase from the *A. fulgidus* genome, named AF-Est2 (locus tag: AF1537). This belongs to the  $\alpha/\beta$  hydrolase 6 family and the X group of the ESTHER classification. This enzyme was biochemically characterised and shown to be highly thermostable and stable in organic solvents and at extremes of pH. The X-ray structure of the AF-Est2 was determined and reveals the presence of a molecule of Coenzyme A (CoA). This is bound in a unique position in the vicinity of the enzyme active site groove, partially obstructing the alcohol binding pocket. AF-Est2 shows a novel utilisation of CoA where the pantetheine moiety could act as a regulatory function for enzyme activity.



## 3.5 Results and Discussion

### 3.5.1 Substrate specificity

The AF-Est2 enzyme was successfully cloned and over-expressed in *Escherichia coli* and purified using a nickel affinity column and size exclusion chromatography. AF-Est2 was tested against a range of *p*-nitrophenyl (*p*NP) esters and was shown to be active against small to medium acyl chain length esters (C2 to C8), which is consistent with its classification as an esterase rather than a lipase. AF-Est2 showed optimum activity against *p*NP-valerate with a  $k_{cat}$  of  $58.9 \pm 0.2 \text{ s}^{-1}$  and a  $K_M$  of  $19 \pm 1 \text{ }\mu\text{M}$  (Table 3-1). The  $k_{cat}$  values are close for the whole range of acyl ester chain lengths and differences in the catalytic efficiency of the enzyme are due to large differences in the  $K_M$  for these different substrates. AF-Est2 was also active against the substrate methyl *p*-toluate which other tested carboxyl esterases are inactive against. Kinetic constants were determined using a phenol red assay giving a  $K_M$  of  $1500 \pm 100 \text{ }\mu\text{M}$  with  $k_{cat}$  of  $1.23 \pm 0.07 \text{ s}^{-1}$ ,  $0.94 \pm 0.02 \text{ s}^{-1}$  and  $0.08 \pm 0.006 \text{ s}^{-1}$  at 70 °C, 50 °C and 30 °C respectively. This shows the enzyme is able to accommodate industrially relevant bulky aromatic carboxylate groups

<b>Table 3-1 – The kinetic characterisation of AF-Est2 using pNP-esters with varying acyl chain length and pNP- benzoate as substrates.</b>			
	$k_{cat} (s^{-1})$	$K_M (\mu M)$	$k_{cat} / K_M (s^{-1} \mu M^{-1})$
C2	$21.5 \pm 0.2$	$620 \pm 10$	$0.0349 \pm 0.0007$
C3	$47.8 \pm 0.5$	$254 \pm 8$	$0.189 \pm 0.006$
C4	$55.5 \pm 0.5$	$92 \pm 3$	$0.61 \pm 0.02$
C5	$58.9 \pm 0.2$	$19 \pm 1$	$3.1 \pm 0.2$
C8	$46.0 \pm 0.3$	$120 \pm 10$	$0.37 \pm 0.03$
C12	0	N/A	0
Benzoate	0	N/A	0
No activity was detected for substrates with an acyl chain length of twelve carbons or more.			

### 3.5.2 Enzyme stability and activity at elevated temperatures

To test the temperature tolerance of AF-Est2, it was incubated at elevated temperatures for 30 min and then cooled to room temperature before being assayed. No loss of activity was observed after incubation at temperatures up to 80 °C. This is close to the optimum growth temperature of *A. fulgidus*. As the incubation temperature was increased past this point the enzyme became progressively denatured, retaining no activity after incubation at 95 °C (Figure 3-1a). The temperature optimum for activity of AF-Est2 was determined by performing the assay at increasing temperatures. The activity increased with temperature following the Arrhenius equation up to a maximum of 80°C. A high level of spontaneous substrate hydrolysis prevented measurement of activity beyond 80 °C.

Activity was also measured whilst incubating AF-Est2 at 70 °C, 50 °C and 30 °C for an extended time period. AF-Est2 was able to retain over 50 % activity for one week at 70 °C showing the enzyme would be an excellent choice for applications where an elevated temperature may be required for solubilisation of industrial substrates or

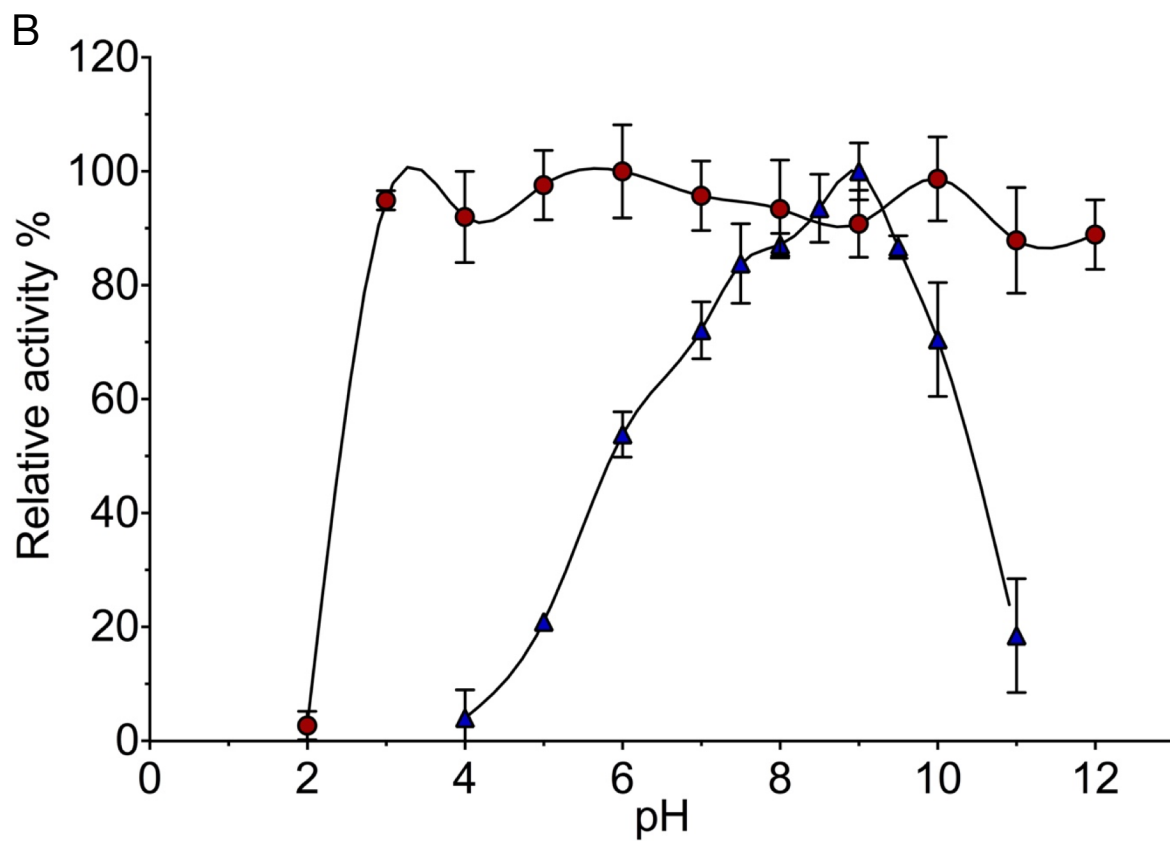
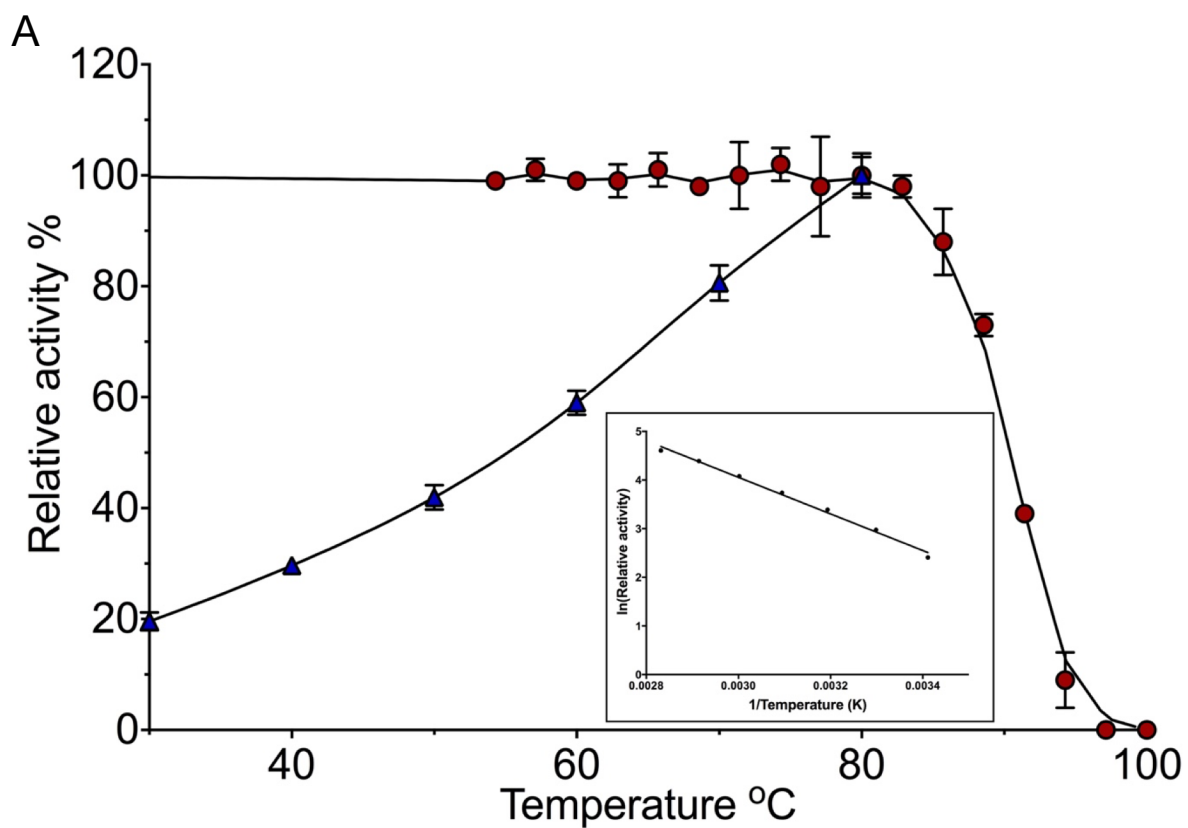
where enzyme stability under reactions conditions is an issue. At 30 °C and 50 °C an excess of 50 % activity remained after 20 days.

### **3.5.3 Enzyme stability and activity at different pHs**

Features adopted by proteins for stability at high temperatures often allow them to withstand other denaturing conditions such as extremes of pH, presence of organic solvents and proteolysis. We therefore tested the activity of AF-Est2 at pH 7.5 after incubation for one hour at a range of different pH values. No significant change in activity was observed for all incubation pH values between 3.0 and 12.0, showing that the enzyme is stable across an unusually wide range of pH values (Figure 3-1b). Incubation at pH 2.0 caused inactivation of the enzyme. Activity assays for AF-Est2 at different pHs showed that the enzyme is active over a broad range of pH values from 5.0 to 11.0 with optimum activity at pH 9.0.

### **3.5.4 Solvent stability**

The residual activity of the AF-Est2 was tested after incubation in a range of common organic solvents. The enzyme retained activity in nearly all of the solvent conditions after incubation for 1 hour in buffer containing 25 mM Tris-HCl pH 7.5, 100 mM NaCl, and either 10 %, 25 % or 50 % of methanol, ethanol, isopropanol, DMSO, acetonitrile and acetone. Only incubation in 50 % isopropanol was significantly detrimental to the enzyme activity resulting in 23 % relative activity when compared to the control. As with the stability of AF-Est2 at extremes of pH, adaptations for thermal stability allow the enzyme to withstand the denaturing effects of a range of organic solvents.



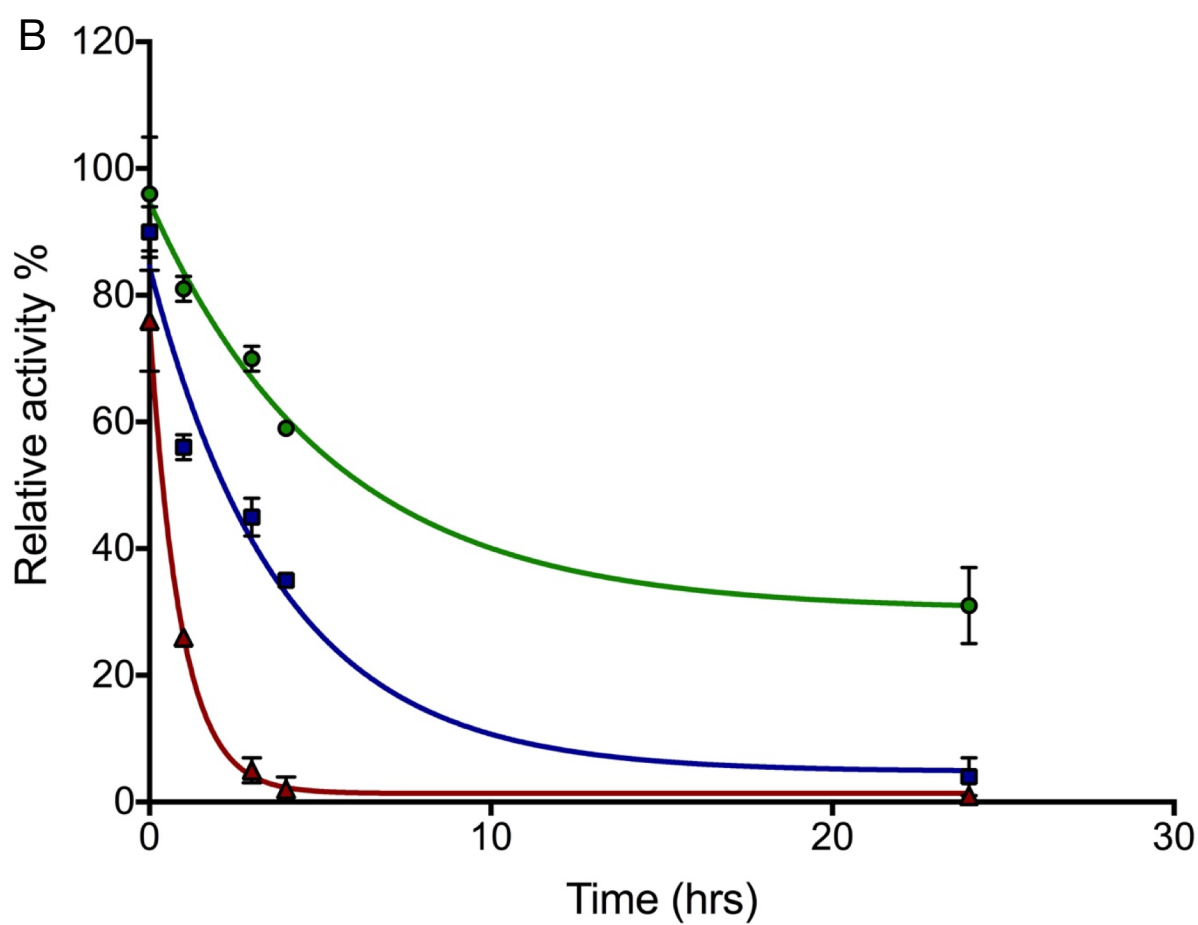
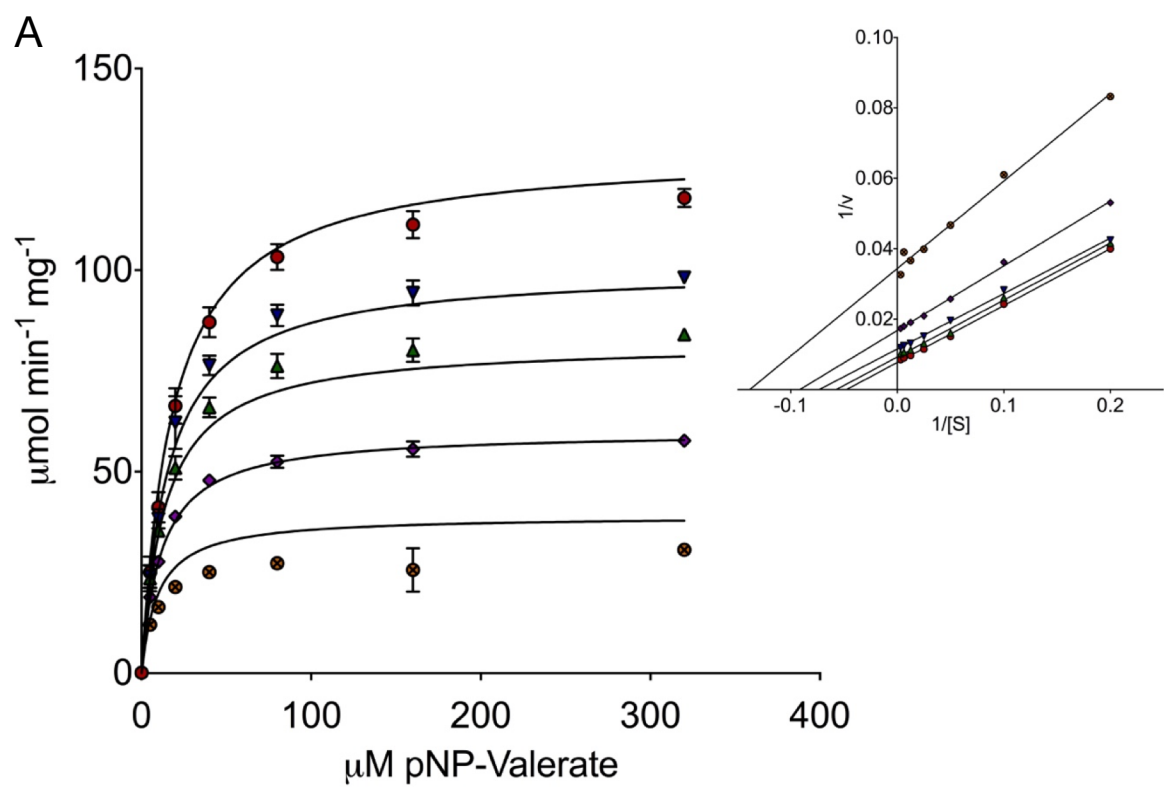
### Figure 3-1 – Effects of temperature and pH on AF-Est2

- a) The effects of temperature on AF-Est2. Blue triangles ▲ show the relative activity of AF-Est2 with increasing temperature, relative to a maximum activity at 80°C. Red circles ● show the relative activity of AF-Est2 incubated for half an hour at various temperatures, against a control kept at room temperature. The insert shows the linearity of the Arrhenius plot of the relative activity with increasing temperature.
- b) The effects of pH on AF-Est2. Blue triangles ▲ show the relative activity of AF-Est2 at various pH values, relative to its maximum activity at pH 9.0. Red circles ● show the activity of AF-Est2 after incubation at various pH values for hour, before assaying under standard conditions.









### 3.5.5 Inhibitors

To assist in determining the mechanism of AF-Est2, the effect of common hydrolase inhibitors on the enzyme was tested. The serine protease inhibitors phenylmethanesulfonyl fluoride (PMSF) and benzamidine and the carboxylesterase inhibitor benzil were used in a range of concentrations up to 2 mM. Both reversible inhibitors benzamidine and benzil showed no effect on the reaction rate of AF-Est2 after incubation for half an hour. To fully inhibit AF-Est2, incubation with 100 µM PMSF for 30 min was required. When PMSF was added during the enzymatic reaction a significantly lower inhibition of the enzyme was observed. This would suggest that the substrates or products compete with the PMSF for binding in the active site of the enzyme.

Investigating the kinetics of the enzyme inhibition with different concentrations of PMSF (Figure 3-2a) revealed the mode of inhibition to be mixed, with a  $K_i$  of  $1.0 \pm 0.3$  µM and an Alpha value of  $0.3 \pm 0.1$ . A small Alpha value suggests PMSF acts more as an uncompetitive inhibitor than a competitive inhibitor, again suggesting an unusual mode of inhibition for this enzyme (Copeland, 2013).



### Figure 3-2 – Inhibition of AF-Est2

- a) Inhibition of AF-Est2 by PMSF. The effect of increasing concentrations of PMSF on the kinetics of AF-Est2 with *p*NP-valerate. Initial rates with varying concentrations of PMSF are displayed as follows: 0  $\mu$ M PMSF by , 0.1  $\mu$ M PMSF by , 0.2  $\mu$ M PMSF by , 0.4  $\mu$ M PMSF by  and 0.8  $\mu$ M PMSF by . The data has been fitted to the mixed model inhibition equation and the following parameters calculated.  $K_{CAT} = 59 \pm 0.7 \text{ s}^{-1}$ ,  $K_M = 18 \pm 1 \text{ }\mu\text{M}$ ,  $K_I = 1.0 \pm 0.3 \text{ }\mu\text{M}$  and  $\alpha = 0.3 \pm 0.1 \text{ }\mu\text{M}$ . A small alpha value suggests PMSF acts more as an uncompetitive inhibitor, which is reflected when the data is shown as a Lineweaver-Burk plot.
- b) Effect of diamide on AF-Est2 activity. The relative activity of AF-Est2 after incubation with various concentrations of diamide over time, shown as 100 mM , 10 mM  and 1 mM . The activity is relative to a control sample with no diamide. Error bars show the standard error of three replicates.

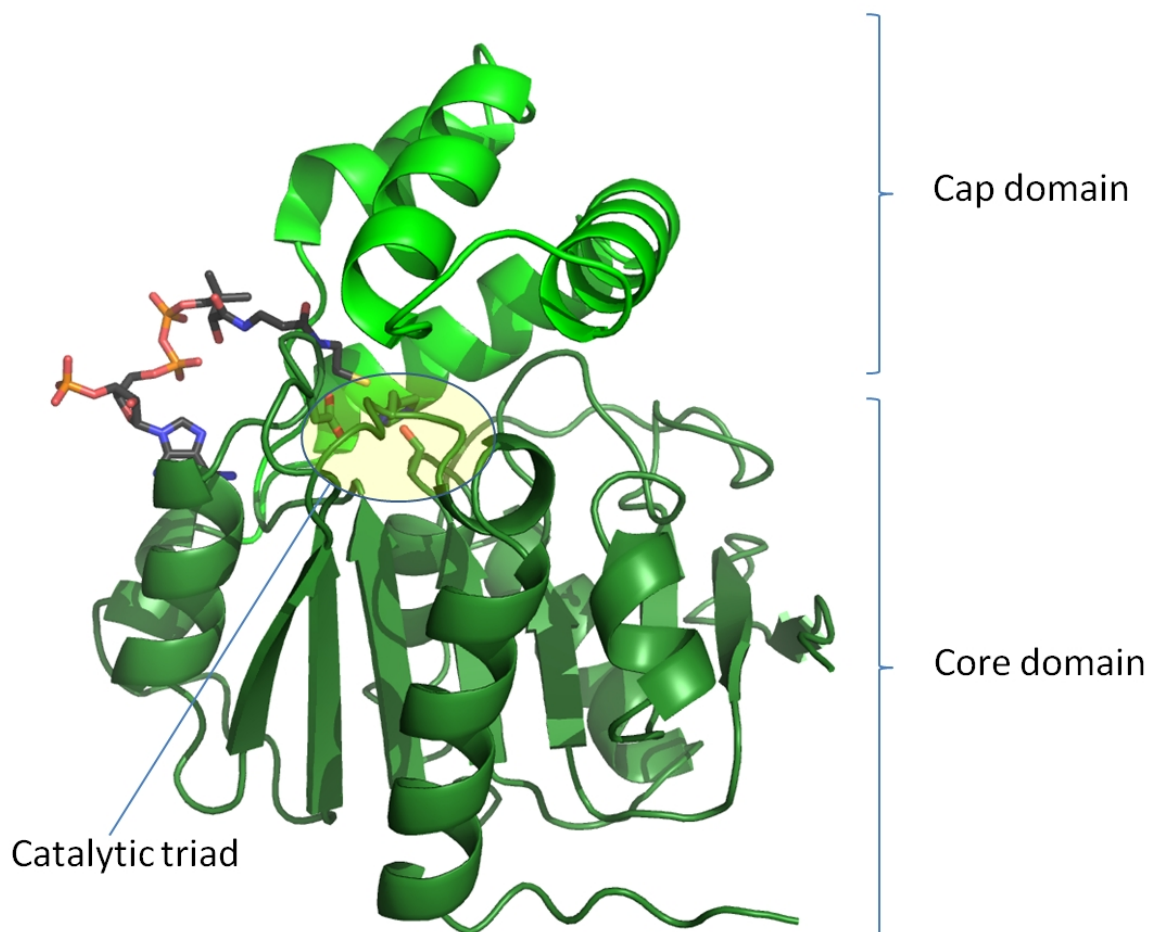
### 3.5.6 Structural determination

AF-Est2 was crystallized in the space group  $P2_12_12_1$  with cell dimensions of  $a=55.6 \text{ }\text{\AA}$ ,  $b=67.6 \text{ }\text{\AA}$ ,  $c=139.5 \text{ }\text{\AA}$ ,  $\alpha=\beta=\gamma=90^\circ$ . The asymmetric unit contains two protein monomers each with a molecular mass of 28.1 kDa giving a solvent content of 46 %. The structure has been isotropically refined to a resolution of 1.4  $\text{\AA}$  with a final  $R_{free}$  of 17.5 (Table 3-2). Several data sets were collected in an attempt to get protein-ligand complexes from crystals soaked or co-crystallised with *p*NP esters, benzyl, benzamidine and PMSF. Data for putative *p*NP-valerate and PMSF complexes were collected to high resolution, but contained no density for the bound ligand.

The structure reveals, as expected, a canonical  $\alpha/\beta$  hydrolase fold core domain (Asp1-Arg117 and Asp176-His254), and a cap domain (Leu118-Phe175) formed by helices  $\alpha 5$ - $\alpha 8$  (Figure 3-3). Six helices of the core domain surround an eight stranded  $\beta$ -sheet of mixed type with connectivity 1,2,-1x,2x,1x,1x,1x with direction ++++++. The catalytic Ser89 residue in both molecules of AF-Est2 is an outlier on the Ramachandran plot. The strained conformation of the catalytic serine residue is observed in the structures of most other  $\alpha/\beta$  hydrolase fold enzymes (Ollis *et al.*, 1992). There are no residues in a *cis* conformation in AF-Est2. The structure also contains two CoA molecules at full occupancy, several ordered PEG molecules and a single citrate molecule bound at a crystal contact. This citrate ligand originates

from the crystallisation solution and could have contributed to the formation of more ordered crystals.

Although the asymmetric unit of AF-Est2 contains two monomers, these do not form oligomers in the crystal. This is in agreement with the chromatography results where the protein eluted as a monomer from a calibrated size exclusion column.



**Figure 3-3 - A cartoon representation of the AF-Est2 structure.**

Showing the core and cap domains in dark green and light green respectively, with the bound CoA molecule shown as a stick model. The catalytic triad residues are shown as stick models and the active site is highlighted as a yellow box. Figure 3-3 to Figure 3-6 were prepared using PyMOL Molecular Graphics System (Schrödinger LLC).



### 3.5.7 Ligand assignment

When most of the AF-Est2 model was built using the original 2.1 Å resolution data, a significant stretch of continuous density became apparent on both the  $2F_o-F_c$  and the  $F_o-F_c$  maps in the proximity of the AF-Est2 catalytic triad in both monomers.

Modelling of a bound polypeptide did not produce a convincing match to the observed density. A molecule of CoA provided the best fit for the un-assigned density. Refinement at the higher resolution of 1.4 Å confirmed the full occupancy of this CoA ligand in the AF-Est2 structure (Figure 3-4a).

### 3.5.8 The CoA binding groove

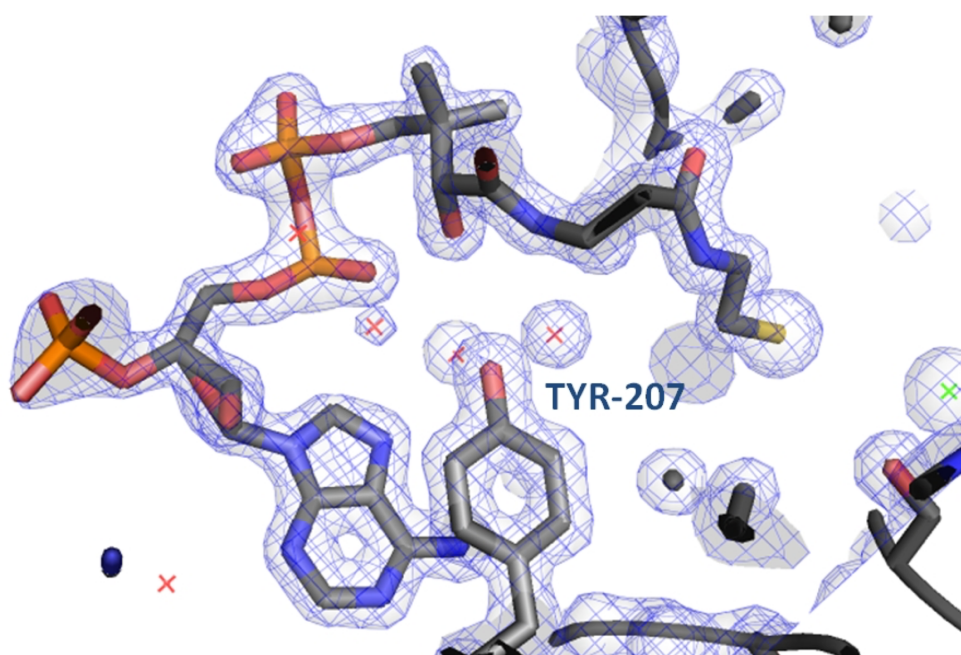
The pantetheine group of the CoA points directly into the active site. The remainder of the CoA molecule is bound in a groove on the surface of the enzyme (Figure 3-4b, c). This groove is formed by the loop region consisting of residues 116-121, preceding  $\alpha 5$  on one side and the loop region of residues 202-207 which precedes  $\alpha 9$  on the other side. There is an ion pair formed by the diphosphate moiety of the CoA and the guanidinium group of Arg117. The residues Arg119 and Arg182 could also form ion pairs with the diphosphate and phosphate groups of CoA, respectively, but do not do so in this structure. The CoA also forms several specific H-bonds and hydrophobic interactions. The adenine ring of CoA is stacked in a pocket between the side chains of Tyr207 and Leu178 (Figure 3-4b) and its N6A atom makes a H-bond to the main chain oxygen of Arg117. The side chain of Lys206 is H-bonded to O2B of the ribose ring of CoA. The N4P and O5P of the pantetheine group of the CoA are H-bonded to main chain oxygen of Leu202 and the main chain nitrogen of Leu121, respectively. It appears that the CoA has a high affinity for the enzyme and is tightly bound since no CoA was added during expression or purification of the enzyme. Extensive dialysis of the purified enzyme did not remove the CoA molecule which may contribute to the high thermostability observed for this enzyme.

The NCBI reference sequence database contains about twenty protein sequences from both archaeal and bacterial sources with higher than 35 % sequence identity to AF-Est2 over more than 90 % of its length. Residues that are involved in CoA binding in AF-Est2 are either conserved or conservatively replaced in the majority of these enzymes. This suggests that these other enzymes could also bind CoA, however, in the absence of their biochemical or structural characterization this

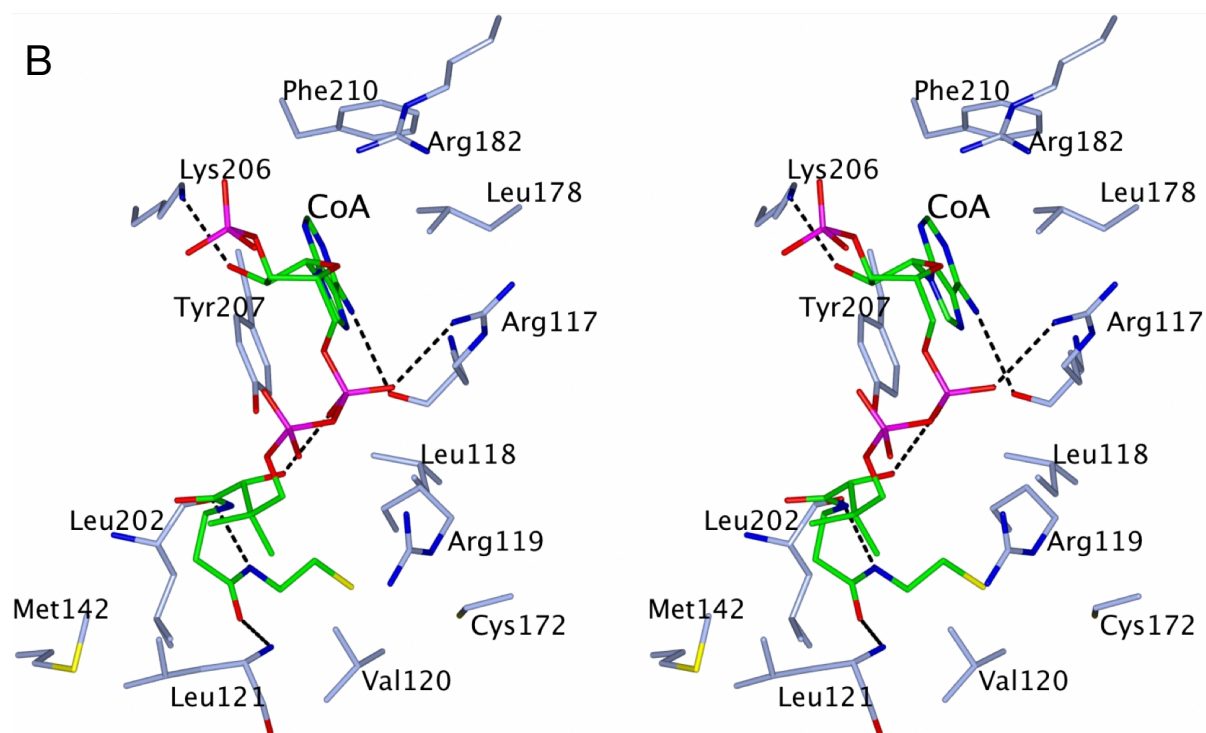
cannot be determined. The structure of the AF-Est2 described in this paper is the first example to show the binding of CoA to an esterase enzyme in this conformation.

Since the AF-Est2 binds CoA this enzyme could have a role in reversible hydrolysis of CoA esters. However, no activity was observed towards the common CoA esters, acetyl-CoA and succinyl-CoA (data not shown) confirming that AF-Est2 does not function as a thioesterase. If CoA esters were a substrate for AF-Est2 they would have moderate affinity for the enzyme to allow the rapid dissociation of the product CoA which would not bind as tightly to the enzyme as observed in these studies. It therefore appears that the tightly bound CoA acts as an integral part of the enzyme.

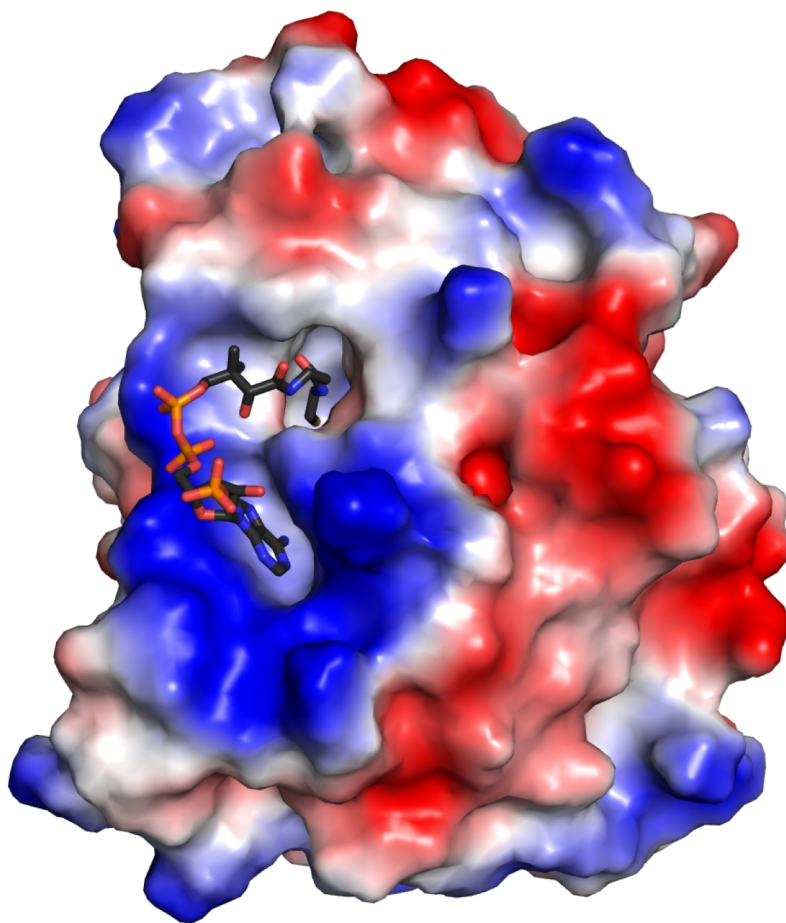
A



B



C



**Figure 3-4 - The interactions of the CoA ligand bound to AF-Est2.**

**a)** A stereo diagram showing the electron density maps in the region of the CoA binding to AF-Est2. The  $2F_o-F_c$  (blue) is contoured at  $1.2 \sigma$  and the  $F_o-F_c$  map is contoured at  $3.5 \sigma$  (green) and  $-3.2 \sigma$  (red). The ligand and amino acid residues are shown as stick models. Solvent molecules are shown as red crosses, a  $Cl^-$  ion is shown as a green cross.

**b)** A stereo diagram showing the CoA binding site of AF-Est2. The CoA molecule is shown as a stick model with carbon atoms coloured in green. Amino acid side chains of residues implicated in the ligand binding are coloured in light blue with hydrogen bonds shown as black dashes. The figure was prepared using CCP4mg (McNicholas *et al.*, 2011).

**c)** The electrostatic potential surface of the AF-Est2 CoA complex. The positive charge is shown in blue and the negative charge is shown in red. The ligand molecule was not used in the surface calculation and is shown as a stick model.

### 3.5.9 Active site

The active site is located at the interface of the core and the cap domains with the conserved catalytic triad composed of Ser89, His228 and Asp200. The catalytic serine is in a strained conformation and is located in a tight nucleophilic elbow at the end of strand  $\beta 5$  with the conserved esterase signature sequence Gly-His-Ser-Leu-Gly.

A comparison of the structure of AF-Est2 with structures of the ligand complexes of other proteins of the  $\alpha/\beta$  hydrolase family 6, allows prediction of its active site pockets. The carboxyl binding pocket of AF-Est2 is likely to be formed by the loop region between  $\beta 3$  and  $\alpha 1$ , and helices  $\alpha 5$ ,  $\alpha 6$  and  $\alpha 7$ . The side chains of residues Met229, Lys154, Phe158, Leu140, Met139 and Ser32, with Phe158 and Lys154, are responsible for defining the size of the carboxyl ester group to be hydrolysed. This largely hydrophobic pocket defines the enzyme's optimal activity towards *p*NP-valerate (Table 3-1). The residues Lys154 and Phe158 (helix  $\alpha 7$ ), at the far end of the carboxyl binding pocket, restrict the chain length of the substrate that can be hydrolysed. It appears that movement of these side chains would be required for the binding of the caprylate chain, which would explain the lower levels of affinity of AF-Est2 towards this *p*NP-ester (Table 3-1). The enzyme's affinity towards the *p*NP-acetate and *p*NP-propionate is also relatively low, which appears to be due to the reduced binding interactions available in the active site for these substrates.

### 3.5.10 CoA obstructs the alcohol binding pocket of the active site

The hydrophobic alcohol pocket is formed by the side chains of residues Leu118, Val120, Leu169 and Cys172, and is partially obstructed by the pantetheine group of the CoA. The alcohol pocket is responsible for binding the *p*NP moiety of the associated substrates. The predicted binding of the *p*NP moiety is thought to be provided by the main chain region of helix  $\alpha 6$  between the side chains of Val120 and Leu169. For successful *p*NP ester hydrolysis, displacement of the pantetheine group of the CoA would be required, since modelling has shown it to be a major steric hindrance for the binding of the relatively large *p*NP group in the alcohol pocket. However high activity of AF-Est2 towards the *p*NP esters suggests that the flexible CoA pantetheine group is displaced from the alcohol pocket upon substrate binding.

### 3.5.11 Thiol oxidation inhibits AF-Est2 activity towards *p*NP esters

The thiol group of CoA is located around 4 Å away from the thiol group of Cys172 in AF-Est2 and would form a disulfide bond with Cys172. The necessity of the CoA pantetheine group displacement for *p*NP ester hydrolysis was confirmed by incubation of the AF-Est2 with the disulfide inducing agent diamide which enforces the formation of a covalent bond between the thiol groups of CoA and Cys172 (Figure 3-4b). This oxidised enzyme shows no activity towards *p*NP-valerate demonstrating that this enzyme is inhibited by the formation of the disulfide. Modelling studies suggest that hydrolysis of substrate esters with small alcohol groups (methyl or ethyl) may not be inhibited by the disulfide formation.

Since the AF-Est2 does not have a signal sequence it is assumed to be a cytosolic protein, with reduced disulfides. It would appear that the CoA ligand may control the levels of activity of AF-Est2 towards carboxyl esters since activity towards esters with larger alcohol groups will be reduced due to the partial obstruction of the active site by the CoA. It is possible that AF-Est2 could have evolved from an ancestral protein with CoA thioesterase activity.

### 3.5.12 Comparison of AF-Est2 with related enzymes of the $\alpha/\beta$ hydrolase family

The AF-Est2 belongs to the  $\alpha/\beta$  hydrolase family 6 in the Pfam classification (Finn *et al.*, 2014). The enzymes of this family are reported to catalyse a variety of different reactions. Examples include a *Pseudomonas fluorescens* esterase (De Yin *et al.*, 2010), bacterial non-heme haloperoxidases (Hofmann *et al.*, 1998), the *Burkholderia xenovorans* 3-oxodipate enol lactonase (Bains *et al.*, 2011) and *Aureobacterium* (-)  $\gamma$ -lactamase (Line, Isupov and Littlechild, 2004). Depending on the features of the active site and the type of reaction that is catalysed, some of the enzymes have esterase activity.

The closest sequence homologues of AF-Est2 in the structural database are the C-C bond hydrolases such as the tetrameric Mhcp (Dunn *et al.*, 2005) from *E. coli* (PDB 1U2E) and the tetrameric meta-cleavage product (MCP) hydrolase (Hsad) from *Mycobacterium tuberculosis* (Lack *et al.*, 2007) (PDB 2VF2). The latter enzyme (with

only 27 % identity to AF-Est2 over 95 % of the amino acid sequence) catalyses the hydrolytic cleavage of a C-C bond in an intermediate of cholesterol metabolism and is also capable of catalysing ester bond hydrolysis *in vitro* (Lack *et al.*, 2007). The active site Ser114Ala mutation of Hsad has allowed elucidation of the conformational changes of the enzyme structure which accompany the substrate binding (Lack *et al.*, 2010) (PDB 2WUF). The relative positions of the cap and core domains differ between Hsad and AF-Est2 which causes a shift of the helix  $\alpha 8$  by 6.5 Å towards the catalytic triad in AF-Est2 (Figure 3-5a). The helices  $\alpha 5$ ,  $\alpha 6$  and  $\alpha 7$  of the cap domain of AF-Est2 are shifted by approximately 3 Å in relation to their equivalent helices in the Hsad structure. The loop region between the core and the cap domain is involved in the CoA binding in AF-Est2. This loop is significantly longer in Hsad and occupies the space of the CoA groove in AF-Est2. The differences of the cap domain positions would prevent the binding of the MCP substrates to AF-Est2. In Hsad, the active site entrance on the interface of the domains is open in the absence of ligand and partially closes in its presence (Lack *et al.*, 2010). However, it is much more closed in AF-Est2, even when the obstructing CoA molecule is not taken into account. The shorter loop region between helices  $\alpha 8$  and  $\alpha 9$  in AF-Est2 would restrict the relative secondary structure movements observed in the Hsad enzyme.

The MCP hydrolases have two conserved sub-sites: a hydrophilic polar sub-site binding a substrate dienoate moiety, and a non-polar sub-site binding the hydrophobic part of the molecule. The different position of helix  $\alpha 8$  and the loops connecting the core and the cap domains would not allow AF-Est2 to form such a non-polar binding site. The polar binding site responsible for MCP hydrolase activity also shows low conservation of both the positioning of the main chain backbone and the amino acid sequence in this region. Therefore, it is not expected that AF-Est2 would have C-C hydrolase activity.

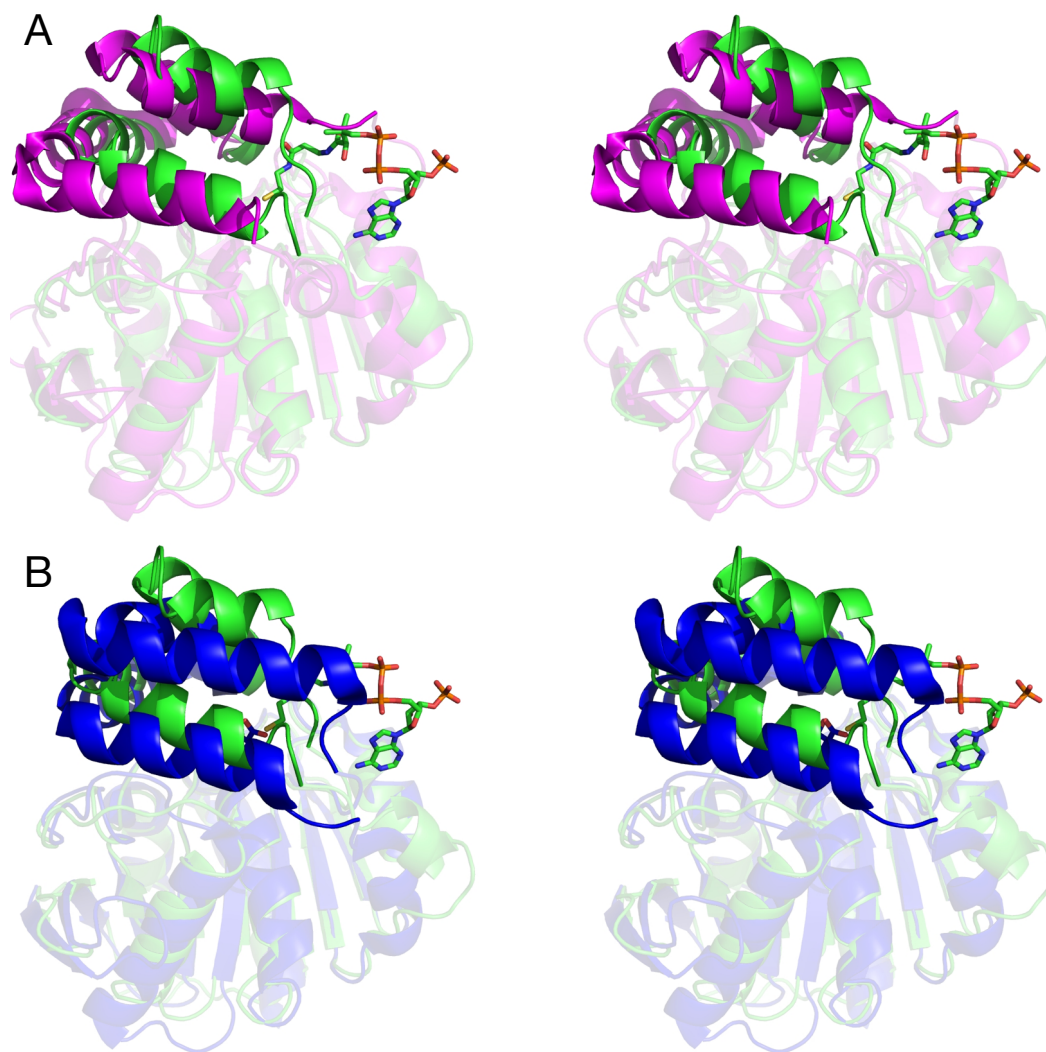
AF-Est2 also shares 27 % identity (over 79 % of the amino acid sequence) with the monomeric *E. coli* BioH carboxyl esterase (Sanishvili *et al.*, 2003) (PDB 1M33), an enzyme involved in biotin biosynthesis. The enzyme has a preference for carboxyl esterase activity towards short acyl chain substrates and has weak thioesterase activity (Sanishvili *et al.*, 2003). Whilst the core domains of the two enzymes superimpose closely, there is a relative movement of the helices  $\alpha 5$ - $\alpha 8$  that comprise the cap domain. The relative movement of the core and the cap domains between

the model and the target could be the reason why an MR solution could not be found for AF-Est2 using Hsad and BioH as models, despite their greater sequence similarity.

AF-Est2 and BioH display little structural conservation in the active site region. One of the two helices in BioH which define the carboxyl pocket ( $\alpha 5$ ) is displaced 5 Å away from the active site in AF-Est2 and the other ( $\alpha 6$ ), which defines the far edge of the pocket, is in a different orientation. BioH shows optimal activity towards *p*NP-acetate and *p*NP-propionate with lower activities observed for *p*NP-butyrate and *p*NP-caprylate. In contrast, AF-Est2 shows a similar activity towards acetate and a much greater activity towards *p*NP-propionate and optimal activity towards *p*NP-valerate. The side chains of Gln147 (equivalent Ala143 in AF-Est2), Trp81 (His88) and Leu24 (Ser32) give BioH a more restrictive carboxyl binding pocket compared to AF-Est2, thereby explaining the difference in substrate specificity between the two enzymes.

Using the DaliLite server (Holm and Rosenstrom, 2010), AF-Est2 shows best structural alignment to an enol lactonase (Bains *et al.*, 2011) (RMSD 2.2 Å; PDB 2XUA) and a carboxyl esterase from a newly isolated thermophilic *Planctomyces* species, *Thermogutta terrifontis* (TtEst; RMSD 2.5 Å) recently described by our group (Sayer *et al.*, 2015). The core domains between TtEst and AF-Est2 superimpose closely and the relative positions of the core and the cap domains are similar (Figure 3-5b). However, helices  $\alpha 5$  and  $\alpha 8$  are displaced in AF-Est2 compared to the TtEst.





**Figure 3-5 - A comparison of the overall fold of AF-Est2 with that of related proteins.**

a) A stereo diagram showing the superposition of AF-Est2 (green) with the open form of Hsad (magenta; PDB 2VF2). The core domains are shown as transparent cartoon models and the cap domains are shown as full colour cartoon models. The AF-Est2 cap domain restrains the size of the ligand capable of binding to the enzyme active site. The CoA ligand is shown as a stick model.

b) A stereo diagram showing the superposition of AF-Est2 (green) with the malate complex of the TtEst (PDB 4UHE). The core domains are shown as transparent cartoon models and the cap domains are shown as full colour cartoon models. The relative positions of the core and cap domains are more similar between AF-Est2 and TtEst, than between AF-Est2 and Hsad. The ligands CoA and the malate which maps the alcohol binding pocket in TtEst are shown as stick models.

The carboxyl binding pocket in AF-Est2 occupies a larger volume than that in TtEst. This is consistent with AF-Est2's preference towards larger butyrate and valerate esters for which the TtEst shows limited activity (Sayer *et al.*, 2015). Mutagenesis studies in TtEst have confirmed that Leu37 hinders the binding of esters with side chains larger than propionate. The corresponding Ser32 in AF-Est2 allows binding of the larger carboxyl side chains. Although most of the residues in the carboxyl pocket are not conserved between AF-Est2 and TtEst, they both are predominantly hydrophobic.

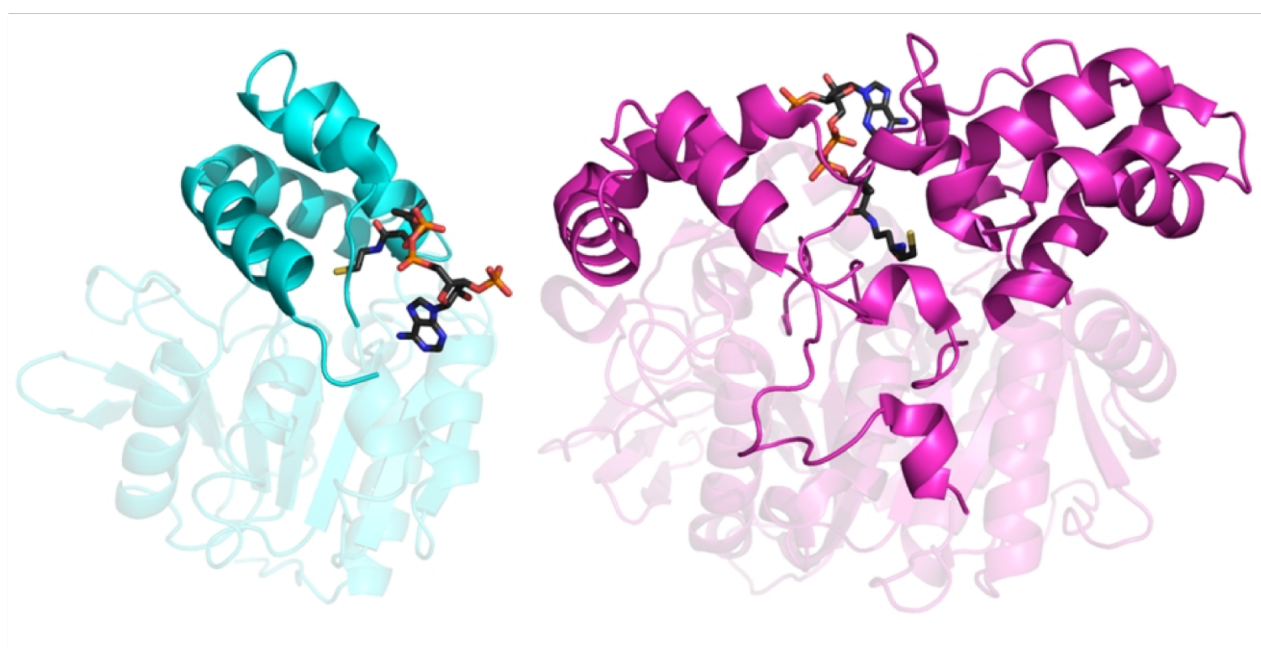
The 3-oxodipate enol lactonase and (-)  $\gamma$ -lactamase, which both are active only towards pNP-acetate, have bulky Trp side chains obstructing the binding of the longer chain substrates in the carboxyl pocket.

The alcohol pocket in AF-Est2 is significantly smaller than that of TtEst and the 3-oxodipate enol lactonase. This is due to obstruction by the panthothine group of the CoA molecule and to the different position of helix  $\alpha 5$ .

### **3.5.13 Comparison of CoA binding with another $\alpha/\beta$ hydrolase fold thioesterase**

Since the results presented in this paper suggest that the AF-Est2 has inherited the CoA binding from an ancestral protein with thioesterase activity, it is interesting to compare this ligand binding in AF-Est2 with other proteins of the  $\alpha/\beta$  hydrolase fold family. Since only a relatively small proportion of CoA thioesterases have the  $\alpha/\beta$  hydrolase fold, a search of the PDB revealed a single enzyme of this family which also binds CoA. This human carboxyl esterase 1 (PDB 2DQZ and 2H7C) is involved in cholesterol ester hydrolysis, fatty acyl CoA hydrolysis, acyl-CoA:cholesterol acyl transfer, and fatty acyl ethyl ester synthesis (Bencharit *et al.*, 2006). The monomer of the hexameric human enzyme is twice as large as AF-Est2; however, the core domains of the two proteins superimpose well. In contrast, the cap domain of the human esterase has no similarity to the cap domain of AF-Est2 and contains three distinct ligand binding sites. The mode of binding of the CoA ligand to one of these sites in the human carboxyl esterase shares little similarity to that observed in AF-Est2 (Figure 3-6). In AF-Est2, the CoA molecule is bound in the groove on the surface between the C-terminal part of the core domain and the cap domain. In the

human carboxyl esterase the CoA binds between the N-terminal part of the core domain and the extended cap domain. This results in the pantetheine groups of CoA coming into the enzyme active sites of AF-Est2 and human carboxyl esterase 1 from opposite directions. These differences confirm that AF-Est2 shows an entirely novel form of CoA binding and suggests that this enzyme has evolved to utilise the pantetheine group of CoA as an integral part of its active site.



**Figure 3-6 - A comparison of the CoA binding to both AF-Est2 (cyan) and human carboxylesterase 1 (magenta) shown as cartoon models in the same orientation.**

The core domains are shown as transparent models and the cap domains are shown in full colour. The CoA ligands bound to the two proteins are shown as ball-and-stick models with carbon atoms coloured in grey. The CoA molecules bind to different parts of the core domain in these two proteins and their pantetheine groups come into the enzyme active site from the different directions.

### **3.5.14 Conclusions**

The AF-Est2 esterase structure has shown a novel binding mode for a CoA molecule which is located in a groove overlapping the active site of the enzyme where the pantetheine moiety is partially obstructing the alcohol binding pocket. This suggests

that the CoA may have a role in reducing activity towards esters with a large alcohol group. The AF-Est2 does not appear to be related to the  $\alpha/\beta$  hydrolase fold eukaryotic carboxyl esterases that also bind CoA. The structures of AF-Est2 and human carboxyl esterase reveal that the pantetheine group approaches the active site from opposite directions between the two enzymes. The AF-Est2 enzyme is so far unique in its mode of binding of CoA and could represent the first example of a new group of esterases. The AF-Est2 reported here shows no thioesterase activity despite the fact that it could have evolved from an ancestral protein with CoA thioesterase activity.

The detailed biochemical and structural comparisons of AF-Est2 with other related Pfam family 6  $\alpha/\beta$  hydrolase enzymes as reported in this paper is of general interest from the aspect of natural enzyme evolution which has occurred within this large enzyme family.

The AF-Est2 esterase from the thermophilic archaeon *Archaeoglobus* has potential applications in industrial biocatalysis due to its overall stability to high temperature, a broad pH range and tolerance to organic solvents.

## 3.6 Methods

### 3.6.1 Expression and purification

The gene encoding AF-Est2 (locus tag: AF1537; Uniprot accession number: O28735) was PCR-amplified, without its stop codon, using chromosomal DNA of *A. fulgidus* as a template and the two primers 5'-  
GCGCCATGGACCTGGAGAGAGTATTCATCG-3' and 5'-  
GGGCTCGAGAACCCCAACTTTTTTGAGAACTTTTCAAGCGC-3', introducing respectively a *Nco*I and *Xho*I restriction site. The generated PCR product was digested by *Nco*I and *Xho*I and the product was purified and ligated into the protein expression vector pET24d (EMD Millipore) digested with the same restriction enzymes, resulting in the plasmid pWUR365 for the expression of the C-terminal 6x-His-tag AF-Est2 protein.

The pWUR365 vector was transformed into the *E. coli* strain BL21-CodonPlus (DE3)-RIPL (Agilent) cells which were grown on Luria-Bertani (LB) medium

containing 50 µg/ml each of kanamycin, chloramphenicol and streptomycin for protein expression. Cultures were grown at 37 °C, 225 rpm until approximately OD<sub>600</sub> 0.6, at which point protein expression was induced by the addition of 1 mM IPTG. The cultures were left shaking at 30 °C for a further 18 hours. Cells were harvested by centrifugation and resuspended in 20 mM Tris-HCl pH 8.0, 10 mM imidazole. Cell lysis was achieved by sonication on ice followed by centrifugation. AF-Est2 was purified using a 1 ml His-Trap FF crude column (GE Healthcare) using an elution gradient from 20 to 500 mM imidazole in 20 mM Tris-HCl pH 8.0, 0.5 M NaCl. The enzyme was then applied to a calibrated Superdex 200 HiLoad 16/60 size exclusion column (GE Healthcare) and was eluted with one column volume of a buffer of 25 mM Tris-HCl, pH 7.5, 0.1 M NaCl.

### 3.6.2 Standard assays

Assays were carried out using *p*NP ester substrates as previously described (Armstrong *et al.*, 1966). A 50 mM stock of each substrate in DMSO was prepared and stored at -80 °C until required. Unless otherwise specified, each reaction was performed in assay buffer containing 50 mM Tris-HCl pH 7.5, 1 mM substrate, 0.125 µg/ml of enzyme and carried out in triplicate in standard 96-well microplates (Greiner 655101). Reactions were started by the addition of substrate after pre-incubation at 30 °C for 5 min, and monitored by following the absorbance at 405 nm over 10 min in a Tecan Infinite M200 PRO plate reader. All activity measurements included a reading for the blank rate of hydrolysis which was subtracted from the observed activity, compensating for any autohydrolysis of the *p*NP esters.

The molar extinction coefficient of *p*NP was determined for every condition prior to measurements. One Unit of esterase activity was defined as the amount of protein releasing 1 µmol/min of *p*NP.

To measure the CoA thioesterase activity a DTNB-based assay (Riddles, Blakeley and Zerner, 1983), which measures the level of free thiols (and hence formed CoA) was performed following incubation of the enzyme with acetyl-CoA and succinyl-CoA.

### 3.6.3 Substrate specificity and kinetics

Kinetic analysis of substrate specificity was carried out by measuring the activity of AF-Est2 against a range of *p*NP esters (C2 to C12 and *p*NP-benzoate) over a range of ten substrate concentrations from 5 to 2500  $\mu$ M. The initial reaction rates were fitted to the Michaelis-Menten equation by non-linear least squares regression using GraphPad Prism v. 5.0.

Phenol red was used to monitor the change in pH as a result of the hydrolysis of the industrial substrate methyl *p*-toluate by the change in absorbance at 540 nm. The 1 ml assay was prepared using 20  $\mu$ g/ml phenol red, 1 mM HEPES pH 8.0 at each assay temperature. A spectrophotometer with a Peltier temperature controller was used for assays at 50°C and 70°C with 13.6  $\mu$ g / ml of enzyme. Assays at 30°C were carried out in a standard 96-well microplate reader in a 200  $\mu$ l volume with 25.0  $\mu$ g / ml of enzyme. All assays were carried out for 10 min after addition of the substrate. A standard curve from 0 to 300  $\mu$ M of *p*-toluic acid was generated for each assay temperature. Activity was measured at eight concentrations of methyl *p*-toluate between 0 and 5000  $\mu$ M and the initial reaction rates were fitted as described above.

### 3.6.4 Effects of temperature

The thermostability of AF-Est2 was assessed by incubating samples at a range of temperatures for half an hour using the temperature gradient function of a T100 thermocycler (Bio-Rad). Following incubation samples were cooled to 4° C before assaying activity against *p*NP-valerate, relative to a control sample kept at room temperature.

To examine the effect of temperature on the activity of AF-Est2, a standard assay buffer was prepared and titrated to the correct pH at temperatures from 20 °C to 80 °C. Using an Evolution 300UV-Vis spectrophotometer (Thermo Scientific) with a Peltier heater, 970  $\mu$ l of assay buffer was equilibrated to temperature in a 1 ml quartz cuvette. The blank rate with 0.5 mM *p*NP-valerate was measured before the addition of 0.125  $\mu$ g of enzyme. A 10 min continuous assay was performed and the data analysed to use only initial linear rates.

To measure the stability of AF-Est2 at 70°C, 50°C and 30°C aliquots of AF-Est2 were prepared at 1 mg/ml in 25 mM HEPES pH 7.5, 100 mM NaCl and 0.01 % sodium azide. Every few days 0.2 µg of enzyme was taken from each aliquot was used to measure activity against 0.5 mM *p*NP-valerate in triplicate, with rates calculated relative to those at the beginning of the experiment.

### **3.6.5 Effects of pH**

To test the stability of AF-Est2 at various pHs, AF-Est2 was incubated for one hour at room temperature in the following buffers covering a range of pH values from 2 to 12: 100 mM KCl-HCl pH 2.0, 100 mM glycine-HCl pH 3.0, 100 mM sodium acetate pH 4.0 and pH 5.0, 100 mM sodium phosphate pH 6.0, 100 mM Tris-HCl pH 7.0, pH 8.0 and pH 9.0, 100 mM glycine-NaOH pH 10.0, 100 mM sodium dihydrogen orthophosphate-NaOH pH 11.0 and pH 12.0. The enzyme was then diluted 250 fold into 50 mM Tris-HCl pH 7.5, to a final concentration of 2.5 µg/ml, and assayed against *p*NP-caprylate (C8).

To access the effect of pH on enzyme activity, reactions were carried out using the buffers described above in place of the standard reaction buffer. The *p*NP-caprylate was used as the substrate to reduce spontaneous hydrolysis at alkaline pHs. Readings were taken at 348 nm, the isobestic point of *p*NP, to remove the effect of pH on the readings.

### **3.6.6 Effect of solvents**

To examine the stability of AF-Est2 in various solvents, the enzyme was incubated for 1 hour in buffer containing 25 mM Tris-HCl pH 7.5, 100 mM NaCl, and either 10 %, 25 % or 50 % solvent. Methanol, ethanol, isopropanol, DMSO, acetonitrile and acetone were tested. A control sample was incubated with no solvent to compare relative activity. Samples were then diluted 1 in 2000 in assay buffer to measure enzyme activity.

### **3.6.7 Effect of inhibitors**

The inhibitors PMSF, benzamidine and benzil at concentrations in the 0 – 2 mM range were incubated with enzyme at 9 nM for 30 min before measuring enzyme activity by the addition of the substrate, 0.5 mM *p*NP-valerate.

The mode of PMSF inhibition was determined using Michaelis-Menten kinetics with eight different *p*NP-valerate concentrations in the presence of five inhibitor concentrations. The best equation to describe the inhibition data was selected by fitting a non-linear least squares regression in GraphPad Prism 5.0.

### **3.6.8 Diamide treatment**

0.25 mg / ml enzyme was incubated in a 1, 10 or 100 mM diamide, 25 mM Tris-HCl pH 7.5, 100 mM NaCl solution for 1 hour before a 1 in 250 dilution into 25mM Tris-HCl pH 7.5, 100 mM NaCl was made. A control incubation was also performed without the addition of diamide. Enzyme assays were performed as specified previously against *p*NP-valerate. Activity was calculated as relative to the control sample.

### **3.6.9 Crystallization**

AF-Est2 was concentrated to ~12 mg/ml using a 10 kDa membrane Vivaspin (Vivaproducts) and microbatch crystallization trials were set up using an Oryx6 crystallization robot (Douglas Instruments) using the JCSG Screen+™ (Molecular Dimensions). The droplet contained a 50:50 ratio of protein solution to screen and was covered with Al's oil (50:50 mix of silicon and paraffin oils) before being stored at 20 °C.

Native crystals appeared within one week in 100 mM ammonium chloride and 10 % (w/v) PEG3350 and were cooled in liquid N<sub>2</sub> using a cryoprotectant consisting of 100 mM ammonium chloride, 10 % (w/v) PEG3350, 35 % (v/v) PEG400, 25 mM Tris pH 7.5 and 100 mM NaCl. Multiple attempts of inhibitor and substrate co-crystallisation produced crystals in conditions related to those of native crystals. Additionally, native crystals were soaked for 30-300 sec in the cryoprotectant containing variable concentrations of ligands.

### **3.6.10 X-ray data collection and structure solution**

Native AF-Est2 data were collected to 2.1 Å and later to 1.4 Å on beamline I03, at the Diamond Synchrotron light source (Didcot, UK) at 100 K in a stream of gaseous nitrogen using a Pilatus detector (Dectris). Data were processed and scaled using XDS(Kabsch, 2010) and AIMLESS(Evans and Murshudov, 2013) in the Xia2(Winter,



Lobley and Prince, 2013) pipeline. All further data and model manipulation was carried out using the CCP4 suite of programs (Winn *et al.*, 2011).

Phases for the initial structure were determined using these 2.1 Å resolution data by the molecular replacement method implemented in the MORDA pipeline (Vagin and Lebedev; <http://www.biomexsolutions.co.uk/morda>). The best solution was found with the bromoperoxidase A1 monomeric model (Hofmann *et al.*, 1998) (PDB 1A8Q) which shares 27 % sequence identity to AF-Est2. Two monomers of the search model were positioned in the AF-Est2 unit cell with a Z-score of 9.6 and in the course of initial refinement  $R_{free}$  was reduced from 57.6 % to 51.9 %. MORDA has estimated the probability of the resulting solution to be correct at 81 %.

The phases were improved by two fold NCS averaging using DM (Cowtan, 2010). The averaged phases were used in phased refinement implemented in the REFMAC5 (Murshudov *et al.*, 2011). The combined partial model and averaged phases produced a good quality map that allowed re-building of AF-Est2 structure in COOT (Emsley *et al.*, 2010). The unknown ligand was identified as CoA at this stage.

Refinement with both anisotropic and isotropic B-factors was attempted since for 1.4 Å the ratio of independent reflections to refined atoms (including alternative conformations) is 20, which is a borderline case for anisotropic B-factor refinement according to PDB REDO (Joosten *et al.*, 2014). After the map inspection isotropic B-factor refinement was selected.

The statistics of the data processing and parameters of the final refined models are given in Table 3-2. The dictionary definitions for *p*NP ester ligands used for docking were prepared using JLIGAND (Lebedev *et al.*, 2012) and the manual docking of ligands was performed in COOT.

<b>Table 3-2 - The AF-Est2 data processing and structural refinement statistics.</b>	
Crystal	COA complex
Beamline (Diamond)	I03
Resolution (Å)	70.29-1.40 (1.44-1.40) <sup>a</sup>
Wavelength (Å)	0.9763
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions	a, b, c= 55.1, 67.2, 140.6 Å; α, β, γ = 90°
No. of protomers in A.U.	2
Solvent content (%); V <sub>M</sub> (Å <sup>3</sup> Da <sup>-1</sup> )	46.4; 2.31
Unique reflections	102188
Redundancy	5.6 (2.5)
Completeness	98.9 (89.6)
<(I)/ σ (I)>	18.9 (2.0)
R <sub>sym</sub> (%)	4.4 (46.1)
Wilson B factor (Å <sup>2</sup> )	24.5
Overall R-factor (%)	15.6
R <sub>free</sub> (5 % total data) %	17.3
Residues refined	508
No. of waters modelled	514
RMSD bond length (Å)	0.011 [0.019] <sup>b</sup>
RMSD bond angle (°)	1.6 [2.0]

Occupancy of ligand	1.0
Average B factor	
Protein (Å <sup>2</sup> )	22.8
Water (Å <sup>2</sup> )	35.5
Ligand	24.8
Ramachandran analysis (% of residues)	
Most favoured	88.7
Additionally allowed	10.9
Generously allowed	0.0
Disallowed	0.4
G-factor	0.1
<p><sup>a</sup> Values for the outer resolution shell are given in parentheses.</p> <p><sup>b</sup> Target values are given in brackets. <math>R_{sym} = \sum_h \sum_j  I_h - I_j(h)  / \sum_h \sum_j I_j(h)</math>, where <math>I(h)</math> is the intensity of the reflections <math>h</math>, <math>\sum_h</math> is the sum over all the reflections and <math>\sum_j</math> is the sum over <math>J</math> measurements of the reflections. The individual atomic B-factors were refined isotropically. <math>R_{cryst} = \sum   F_o  -  F_c   / \sum  F_o </math>. Wilson B-factor was estimated by SFCHECK(Vaguine, Richelle and Wodak, 1999). The Ramachandran plot analysis and G-factor calculation were performed by PROCHECK (Laskowski <i>et al.</i>, 1993).</p>	

### 3.6.11 Accession codes

The coordinates and structure factors for AF-Est2 have been deposited in the Protein Data Bank with code: 5FRD.

## **3.7 Acknowledgments and Other Information**

### **3.7.1 Acknowledgements**

This work was supported by the Hotzyme project (grant agreement no. 265933) financed by the European Union 7th Framework Programme FP7/2007-2013. WF is funded by a BBSRC PhD studentship. MI would like to thank the BBSRC funded ERA-IB grant BB/L002035/1 and the University of Exeter for support. The authors would like to thank the Diamond Synchrotron Light Source for access to beamline I03 (proposals No. MX8889 and No. MX11945) and the beamline scientists for assistance. The work of ML was funded by the Graduate School VLAG Wageningen, the Netherlands.

### **3.7.2 Author contributions**

SVWK and ML identified and cloned the gene within the group of JvdO at Wageningen University. CS over-expressed and crystallised the protein. The protein was biochemically characterised by CS, WF and NH. The structure was determined by CS and MNI. JAL coordinated the project and wrote the manuscript with input from all other authors.

### **3.7.3 Competing interests**

There are no competing financial or other competing interests.

# **Chapter 4 - Characterization of carboxylic acid reductases as enzymes in the toolbox for synthetic chemistry**

## 4.1 Authors

William Finnigan<sup>[a]</sup>, Adam Thomas<sup>[a]</sup>, Holly Cromar<sup>[a]</sup>, Ben Gough<sup>[a]</sup>, Radka Snajdrova<sup>[b]</sup>, Joseph P. Adams<sup>[b]</sup>, Jennifer A. Littlechild<sup>[a]</sup>, Nicholas J. Harmer<sup>\*[a]</sup>

[a] W. Finnigan, A. Thomas, H. Cromar, B. Gough, J.A. Littlechild, N. J. Harmer  
Biosciences, College of Life and Environmental Sciences, University of Exeter,  
Stocker Road, Devon, Exeter EX4 4QD, UK

[b] R. Snajdrova, J. Adams  
Synthetic Chemistry, AC - API Chem – UK, GlaxoSmithKline R&D Ltd, Medicines  
Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK

## 4.2 Preface for inclusion in this thesis

This chapter is made up of the reformatted manuscript for a paper initially submitted for publication in ChemCatChem in May 2016, and again in September 2016 after responding to reviewer comments. mpCAR is used to fill the second step in the enzymatic cascade reaction investigated in subsequent chapters, shown in Figure 1-8. It was chosen for the cascade due to its superior thermostability allowing better operational stability and with the hope of expression in *T. thermophilus*.

All biochemical work was performed by William Finnigan, with the exception of the kinetic characterisation of msCAR which was performed by Holly Cromar. Ben Gough also carried out preliminary work on niCAR and tpCAR. Phylogenetic analysis was performed by Adam Thomas.

All text and figures relating to the phylogenetic analysis of the CAR enzymes, including discussions on similarities with members of the ANL superfamily, was written by Adam Thomas, and edited by all authors.

All other text and figures were written by William Finnigan with editing and contributions by Nicholas Harmer. Final edits were carried out by all authors.

The manuscript has been formatted to match the formatting of this thesis. References have been combined with those in the other chapters of this thesis for a final reference chapter. Figure and table numbers have also been edited for inclusion in the thesis.

This manuscript was accepted by ChemCatChem in November 2016, after responding to further reviewer comments. DOI: 10.1002/cctc.201601249. The accepted version has only minor differences to this thesis version.

## 4.3 Abstract

Carboxylic acid reductase enzymes (CARs) meet the demand in synthetic chemistry for a green and regio-specific route to aldehydes from their respective carboxylic acids. However, relatively few of these enzymes have been investigated. Sequence alignments with members of the ANL superfamily of enzymes shed light on CAR functional dynamics. From a phylogenetic analysis of known and hypothetical CARs, four unstudied enzymes were selected, and for the first time, a thorough biochemical characterization carried out. Kinetic analysis of these enzymes with various substrates shows they have a broad but similar substrate specificity. Electron rich acids are favored, suggesting that the first step in the proposed reaction mechanism, attack by the carboxylate on the  $\alpha$ -phosphate of ATP, is the step determining substrate specificity and reaction kinetics. Effects of pH and temperature provide a clear operational window for these enzymes, while investigation of product inhibition by NADP<sup>+</sup>, AMP and pyrophosphate (PP<sub>i</sub>) indicates that binding of substrates at the adenylation domain is ordered with ATP binding first. This paper consolidates CARs as important and exciting enzymes in the toolbox for sustainable chemistry, providing specifications for their use as a biocatalyst.



## 4.4 Introduction

The demand for 'green chemistry' is of increasing global importance, driven by the need to balance sustainable and efficient resource utilization with the demands and increasing consumption of a rising population (Zhang, 2015). Biological solutions to chemistry challenges are a critical component in meeting this demand. The use of isolated enzymes and cell-based systems that produce negligible dangerous waste, often with higher yields, offers an alternative to traditional chemical processes. In some cases, biological alternatives are more rapid and cost effective than their chemical counterpart (Littlechild, 2011). Despite these potential advantages, enzymes are still under-used in chemistry. Expanding the toolbox of available enzymes is essential for the successful development of new synthetic routes and sustainable manufacturing processes (Meyer *et al.*, 2013).

An important opportunity that is ripe for exploitation is synthetic routes based on organic acids. These compounds have a long history of production by fermentation (Kertes, King and Blanch, 2009). Indeed, multiple carboxylic acids were identified to be "Top Value Added Chemicals From Biomass" (Erickson, Nelson and Winters, 2012), many of which are now being produced industrially. Reduced products of these organic acids, especially optically pure aldehydes and alcohols, are essential building blocks for use in the chemical, pharmaceutical and food industries (He *et al.*, 2004). However, chemical methods for the reduction of carboxylic acids are limited, and require chemicals such as lithium aluminium hydride and sodium borohydride in stoichiometric amounts (Napora-Wijata, Strohmeier and Winkler, 2014).

Two enzyme classes are capable of reduction of organic acids to aldehydes, and a review of the biocatalytic reductions possible by organisms harboring them published (Napora-Wijata, Strohmeier and Winkler, 2014). The aldehyde oxidoreductases (AORs) reversibly oxidize organic aldehydes to their respective acids. The oxidized product is more thermodynamically favorable, and so the equilibrium tends towards this product. AORs are therefore more useful for syntheses that require the oxidation of aldehydes (Heider, Ma and Adams, 1995; Hollmann *et al.*, 2012). In contrast, the carboxylic acid reductases (CARs) catalyze the reduction of a carboxylic acid to an aldehyde at the expense of ATP and NADPH, producing AMP, PP<sub>i</sub> and NADP<sup>+</sup> as by-products (He *et al.*, 2004). The reduction of carboxylic acids into aldehydes by

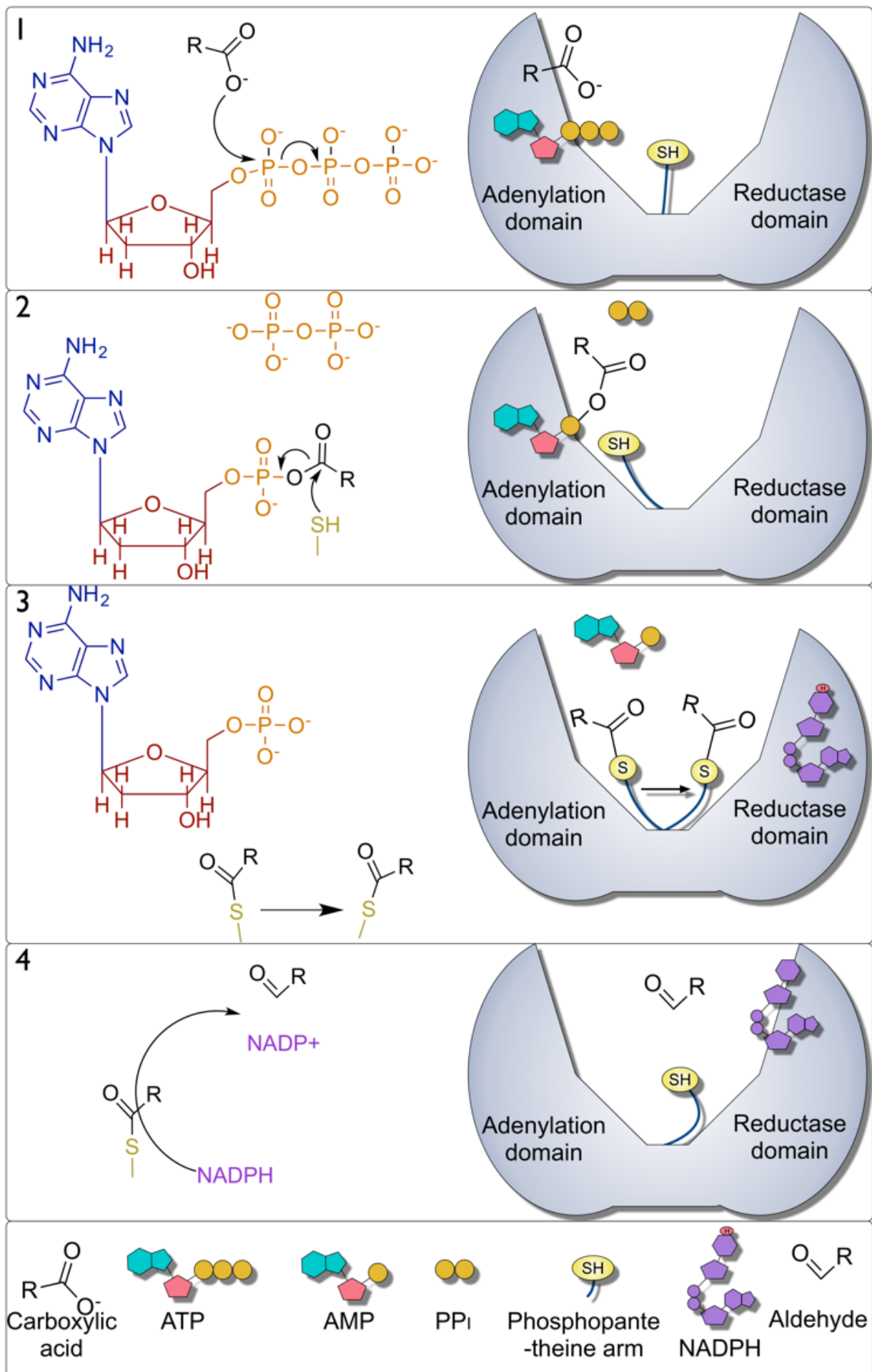
CARs has been confirmed by a number of studies previously, using GC-MS analysis. Products other than the aldehyde have not been detected (Akhtar, Turner and Jones, 2013; Moura *et al.*, 2015). The hydrolysis of ATP makes the reduction of acids to aldehydes by CARs strongly thermodynamically favorable, making them an attractive green route to optically pure aldehydes (Napora-Wijata, Strohmeier and Winkler, 2014). This synthesis can be coupled to other enzymes such as an alcohol dehydrogenase which can provide a complete route to the alcohol product (Akhtar, Turner and Jones, 2013).

Indeed, CARs have been employed in a number of synthetic pathways. These include the production of the flavor vanillin by yeast (Hansen *et al.*, 2009), and a synthetic pathway for the production of propane in *Escherichia coli* (Kallio *et al.*, 2014). These examples both highlight the potential of CARs as part of a toolbox for synthesis of fine chemicals from non-oil-based chemical precursors (Erickson, Nelson and Winters, 2012).

CARs are relatively large, multidomain enzymes of around 130 kDa. They feature an N-terminal adenylation domain, a C-terminal thioester reductase domain that likely adopts a Rossmann fold, and a central phosphopantetheine binding domain (Figure 4-1) (Marchler-Bauer *et al.*, 2015). A phosphopantetheine arm must be covalently attached to a conserved serine in this central domain through the action of a phosphopantetheine transferase for the production of an active enzyme (Venkitasubramanian, Daniels and Rosazza, 2007). Fungal  $\alpha$ -aminoadipate reductases, which are responsible for the reduction of  $\alpha$ -aminoadipate to  $\alpha$ -aminoadipate semialdehyde in lysine biosynthesis, share this domain architecture, also with the requirement for the loading of a central phosphopantetheine prosthetic group (Ehmann, Gehring and Walsh, 1999). However these enzymes have been shown to have a different substrate specificity from CARs (Moura *et al.*, 2015).

Phosphopantetheine arms are most commonly associated with acyl carrier proteins where they maintain an acyl chain in a energetically active thioester bond, with the length and flexibility of the arm allowing access to spatially distinct active sites (Chan and Vogel, 2010). In CARs, the phosphopantetheine arm is believed to act in much the same way, shuttling an attached acyl chain between the N- and C-terminal domains (Venkitasubramanian, Daniels and Rosazza, 2007).

The proposed mechanism of CAR enzymes has four main steps (labelled 1 to 4; Figure 4-1). In the first two steps, the relatively unreactive carboxylic acid is activated to form a thioester with the phosphopantetheine arm at the N-terminal adenylation domain, in a mechanism possibly similar to that of the ANL superfamily of adenyating enzymes such as long chain fatty acid CoA ligases (Hisanaga *et al.*, 2004; Gulick, 2009). (1) ATP and a carboxylic acid enter the active site of the adenylation domain where the  $\alpha$ -phosphate of ATP is attacked by an oxygen from the carboxylic acid, forming an AMP-acyl phosphoester with the release of pyrophosphate (Hisanaga *et al.*, 2004). (2) The thiol group of the phosphopantetheine arm can then nucleophilically attack the carbonyl carbon of the AMP-acyl phosphoester intermediate, releasing AMP and forming an acyl thioester with the phosphopantetheine arm. (3) The phosphopantetheine arm transfers to the C-terminal reductase domain (4) where the thioester is reduced by NADPH, releasing the aldehyde and  $\text{NADP}^+$ , and regenerating the thiol of the phosphopantetheine arm in the process (Napora-Wijata, Strohmeier and Winkler, 2014).



**Figure 4-1 – Proposed mechanism of CAR enzymes.**

- 1:** ATP and a carboxylic acid enter the adenylation domain where a phosphoester intermediate is formed releasing pyrophosphate in the process.
- 2:** the thiol of the phosphopantetheine arm nucleophilically attacks the carbonyl carbon of this intermediate forming a thioester intermediate with the phosphopantetheine arm, releasing AMP.
- 3:** the phosphopantetheine arm transfers to the reduction domain where,
- 4:** the thioester bond is reduced by NADPH releasing an aldehyde product, regenerating the phosphopantetheine thiol group and producing  $\text{NADP}^+$ .

Relatively few CARs have been explored to date. CARs were first described in *Neurospora crassa* as an aryl-aldehyde: NADP<sup>+</sup> oxidoreductase (Gross and Zenk, 1969). Subsequently, CARs were characterized from *Nocardia asteroides* JCM 3016 (Kato *et al.*, 1991) and later *Nocardia iowensis* (Li and Rosazza, 1997) (referred to as niCAR here) when they were reclassified as carboxylic acid reductases (Li and Rosazza, 1997). Characterization of the *Nocardia asteroides* JCM 3016 CAR was performed by comparing the relative activity of this enzyme towards various aromatic substrates (1 mM concentration). This CAR was reported to prefer 3-substituted benzoates and aliphatic acids that were substituted with a phenyl group. No reaction of this CAR with simple aliphatic acids was reported. The optimum pH for activity of this enzyme was pH 7.5, and the optimum temperature for activity was 40 °C (Kato *et al.*, 1991).

The relative activities of the *Nocardia iowensis* CAR (niCAR) against various aromatic substrates have also been reported. The highest activity was achieved with indole-5-carboxylic acid, which was the most activated carboxylic acid tested. Substrates with 2-substituted benzoates or ring-deactivating groups showed no or very low levels of activity. The reduction of racemic ibuprofen by whole *Nocardia iowensis* cells gave a enantiomeric excess (ee) of 61.2 %, which has been attributed to enantio-selectivity by niCAR based on kinetic data for its reduction of (S)-(+)-Ibuprofen and (R)-(-)-Ibuprofen enantiomers (Li and Rosazza, 1997). The requirement for the presence of a phosphopantetheine transferase for the loading of a phosphopantetheine group onto the CAR enzyme was shown for niCAR and is presumed to be the case for all the CAR enzymes (Li and Rosazza, 1997).

A CAR from *Mycobacterium marinum* has also been described and its application for the reduction of fatty acids to fatty alcohols explored. This CAR is active against fatty acids between two and eighteen carbons in length (Akhtar, Turner and Jones, 2013). CAR activity has also been reported in the zygomycete fungus *Syncephalastrum racemosum*, although a responsible gene has yet to be identified (Brenna *et al.*, 2015).

Recently a characterization of CARs from *Nocardia iowensis*, *Nocardia brasiliensis*, *Mycobacterium marinum* and *Mycobacterium smegmatis* showed CARs to prefer substrates where the carboxylic acid was the only polar or charged group, giving a

useful insight into the substrate specificity of these enzymes. Also, a model was developed for the prediction CAR reactivity using this and previous CAR data (Moura *et al.*, 2015). It is worth noting that the msCAR characterized by Moura *et al.* is distinct from msCAR used in this study.

Here, we have produced a detailed phylogeny of the CARs and identified four previously undescribed CARs for further study that are broadly spread across this phylogeny. With the addition of niCAR for comparison to earlier work, a thorough biochemical characterization was carried out on each. We investigated the effects of temperature and pH to identify suitable conditions for the use of CARs in biocatalytic reactions. We further performed a full kinetic analysis on a range of aromatic and aliphatic substrates with these CARs to look for potential differences in their substrate specificity and to examine the effects of various functional groups on their kinetic parameters. Finally, we describe potential issues of product inhibition with the CAR enzymes. Our investigation provides a more thorough description of the factors to be considered when using the CAR enzymes in industrial biocatalysis.

## **4.5 Results**

### **4.5.1 Alignment and phylogenetic analysis**

CAR adenylation domains were aligned with a firefly luciferase, a fatty acyl-CoA ligase and a reductase domain from a non-ribosomal peptide synthetase, all from the ANL superfamily (Supplementary Figure 4-1). CARs share ~20% sequence homology with other ANL superfamily members. Members of the ANL super family catalyze the initial adenylation of a carboxylic acid to form an acyl-AMP intermediate, which is generally followed by the formation of a thioester. The family name is based on three of its sub families: Acyl-CoA synthetases, the non-ribosomal peptide synthase (NRPS) adenylation domain, and the Luciferase enzyme (Gulick, 2009). Previous alignment and crystallography studies have identified nine motifs that are conserved within the superfamily. Of the nine, five are strongly conserved within the CARs, including the active site ppxTSGSTGxPK, rGxTE and TGD motifs (where p=aliphatic and r=aromatic residues). These motifs are considered “signature” to the ANL superfamily, and are involved in the hydrolysis of ATP (Marahiel, Stachelhaus

and Mootz, 1997). The remaining four motifs are also present albeit with lower conservation.

A total of 48 unique sequences showing homology to known CAR proteins were gathered using pBLAST, or mined directly from GenBank by raw text searches (Supplementary Figure 4-2). All sequences identified were solely from Subclass *actinobacteridae*. Within this Subclass, sequences were obtained from families *Streptomycetaceae* and *Corynebacterinae*.

A masked multiple sequence alignment of the dataset was shown to be best fit to the Whelan and Goldman model of amino acid substitution, with a discrete gamma distribution of mutation rates and an assumed presence of invariant sites (WAG+I+G). This model was implemented into a Bayesian phylogenetic reconstruction (Figure 4-2). According to 16S data, the *Streptomycetaceae* are thought to have evolved before the *Corynebacterinae*. However, rooting the tree on the streptomycetes has poor parsimony as numerous gene loss events would have had to have occurred for this to be the case (Lodders and Kämpfer, 2007). Instead, due to an outgroup being unobtainable for this dataset, we opted to root the tree on its midpoint. The tree is extremely well supported, with all nodes possessing a confidence score of >0.75, and only four of forty-six biologically relevant nodes being scored at below the highest possible confidence score of 1.

In order to better understand how CAR functionality differs across clades, we selected sequences for characterization from a range of host organisms that broadly cover distinct areas of the phylogenetic tree.



<b>Table 4-1– Carboxylic acid reductases chosen for this study.</b>		
<b>Abbreviation</b>	<b>Source</b>	<b>NCBI Reference:</b>
mpCAR	<i>Mycobacterium phlei</i>	WP_003889896.1
msCAR	<i>Mycobacterium smegmatis</i>	AFP42026.1
niCAR	<i>Nocardia iowensis</i>	Q6RKB1.1
noCAR	<i>Nocardia otitidiscaviarum</i>	WP_029928026.1
tpCAR	<i>Tsukamurella paurometabola</i>	WP_013126039.1
<p>Five carboxylic acid reductases were chosen for a through biochemical characterization from a range of host organisms containing putative CAR's. CAR abbreviations have been chosen to reflect their source. NCBI ascension numbers are shown which may be used to access the protein sequences.</p>		



### 4.5.2 Expression and Purification

CAR enzymes (Table 4-1) were expressed in *E. coli* and purified from the cell lysates by nickel affinity chromatography followed by gel filtration in order to obtain a high level of purity (Supplementary Figure 4-3). The optimum conditions for the expression of *Mycobacterium phlei* CAR (mpCAR) in *E. coli* in LB media were determined to be induction at OD<sub>600</sub> 0.6 with 150  $\mu$ M IPTG, followed by incubation for approximately 18 hours at 20 °C, with orbital shaking at 225 rpm (data not shown). Similar conditions were assumed to be suitable for the expression of the other CAR enzymes and indeed all CARs were well expressed. CARs were co-expressed with the Sfp phosphopantetheinyl transferase from *Bacillus subtilis* on a separate plasmid, as the loading of a phosphopantetheine group onto CAR enzymes has been shown to be essential for activity (Venkitasubramanian, Daniels and Rosazza, 2007).

### 4.5.3 Kinetic characterization of CAR enzymes

The CAR enzymes were characterized in terms of their substrate specificity towards a range of aromatic carboxylic acids, a range of aliphatic unsaturated fatty acids, and the cofactors ATP and NADPH. A list of the substrates with their chemical structure can be found in Figure 4-3 and Supplementary Figure 4-4. For each CAR, an initial assay at high substrate concentration (5 mM) was carried out to identify compounds for which CAR had activity. For those compounds where activity was detected, a full kinetic analysis was performed (Supplementary Figure 4-5 to Supplementary Figure 4-9). All the CARs that were tested showed similar  $K_M$  values for NADPH and ATP. For NADPH the  $K_M$  was between 24 and 36  $\mu$ M, whilst for ATP  $K_M$  values of between 64 and 84  $\mu$ M were observed. These values are both well within the physiological ranges for these cofactors and in good agreement with previous studies. (Li and Rosazza, 1997; He *et al.*, 2004; Akhtar, Turner and Jones, 2013) Production of benzaldehyde and 4-methylbenzaldehyde from the derivative acids was confirmed by HPLC, with no other products observed. NADPH consumption was also confirmed as a good measure of aldehyde production (Supplementary Figure 4-10 and Supplementary Figure 4-11).

#### 4.5.4 Investigating the effects of electronic density on aryl substituted carboxylic acid substrates

All of the enzymes that we tested showed strong activity against the classical CAR substrate benzoic acid (compound **1**, Figure 4-3A, Table 4-2), which all previously studied CARs have shown activity against. (He *et al.*, 2004; Akhtar, Turner and Jones, 2013; Moura *et al.*, 2015) A series of substituents of varying electronic configuration were tested (compounds **2-5**, Figure 4-3A, Table 4-2). Compounds with more electron rich systems generally lowered  $K_M$  giving increased catalytic efficiency compared to benzoic acid. Minimal activity was detected with 2-methoxy benzoic acid.

In contrast, compounds that incorporated an electron withdrawing nitro group in the benzene ring (compounds **6-8**, Figure 4-3B, Table 4-2) resulted in a large decrease in the turnover number of the CARs, in most cases inhibiting activity all together. Again, there was no detectable activity with a nitro group in the 2 position; whilst in the *para* position only tpCAR showed a low level of activity. However, all the CARs tested were active against 3-nitrobenzoic acid but with a lower  $k_{cat}$  than benzoic acid. Absorbance at OD<sub>340nm</sub> by nitro compounds was shown not to interfere with the assay (Supplementary Figure 4-12)

#### 4.5.5 Investigating the effect of the aromatic unit on catalytic activity

3-phenylpropionic acid (compound **9**, Figure 4-3C, Table 4-3) has a carboxylate group out of conjugation from the aryl ring, extended away from the aryl ring by two carbons giving the carboxylate group greater flexibility. When tested with the CARs, this change caused a reduction in  $K_M$ , with a similar or slightly lower  $k_{cat}$ . (*E*)-3-phenylprop-2-enoic acid (cinnamic acid; compound **10**, Figure 4-3C, Table 4-3), being a conjugated system, was expected to have activity between **9** and the model compound **1**, benzoic acid. The CAR activity against (*E*)-3-phenylprop-2-enoic acid showed a substantial reduction in  $k_{cat}$  compared to 3-phenylpropionic acid or benzoic acid, with a slight further reduction in  $K_M$ . The cognate compound with a triple bond (phenylpropynoic acid; compound **11**, Figure 4-3C, Table 4-3) showed very low or no detectable activity in the CAR reaction.

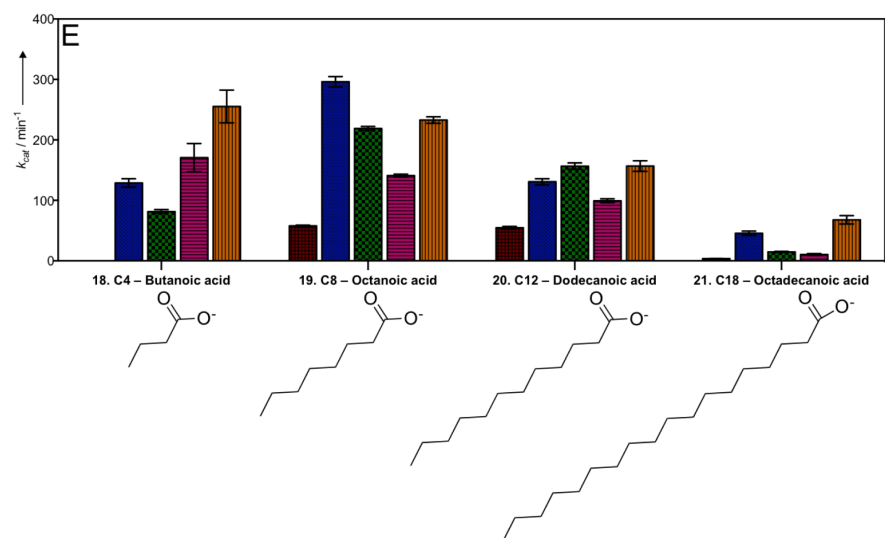
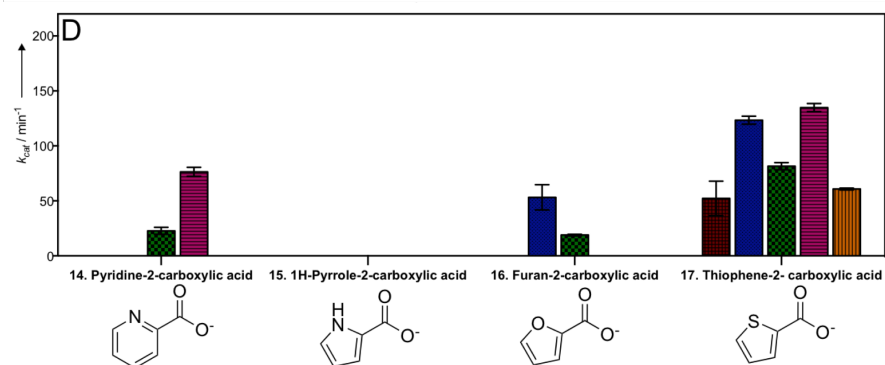
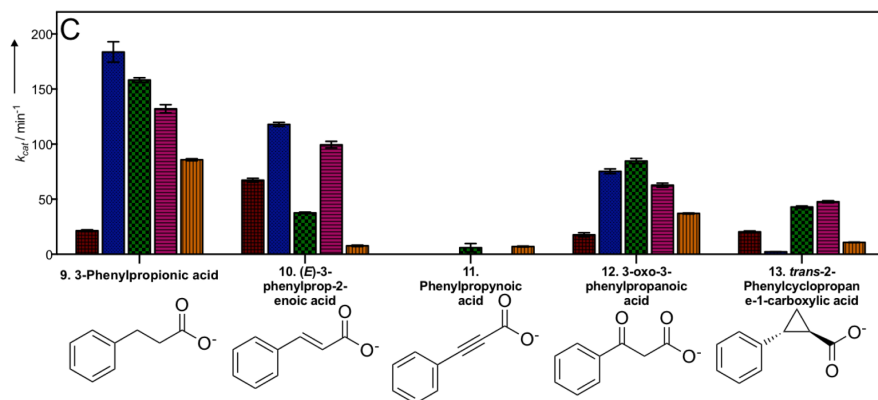
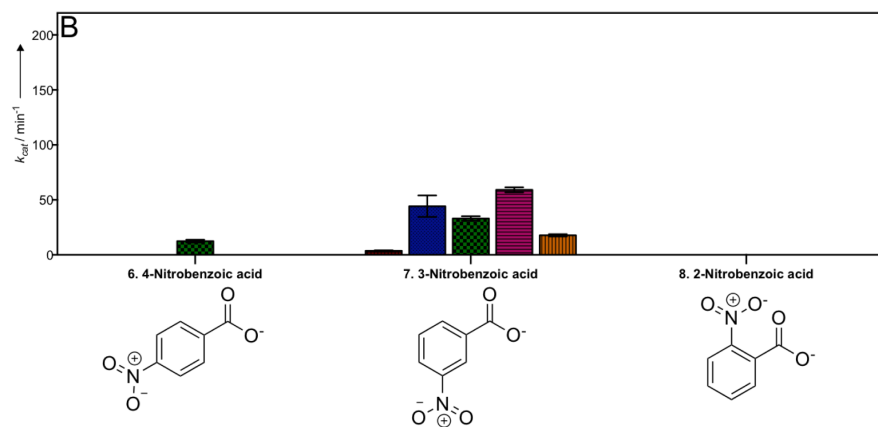
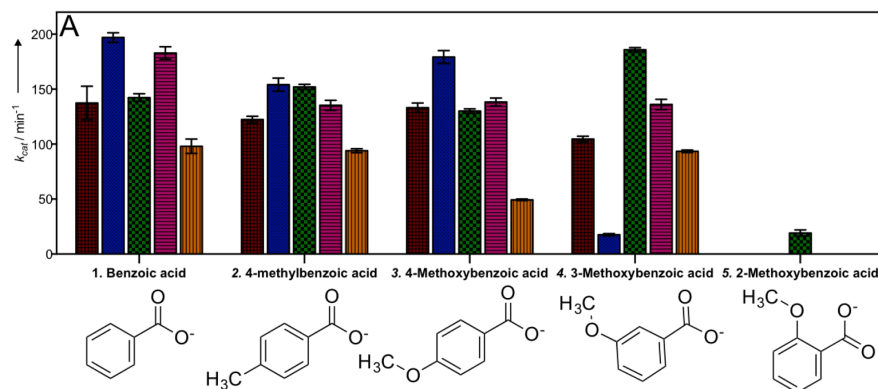
Two other compounds were tested: firstly, the  $\beta$ -keto acid 3-oxo-3-phenylpropanoic acid (compound **12**, Figure 4-3C, Table 4-3) showed an increase in  $K_M$  with mpCAR, msCAR and tpCAR, but a decrease in  $K_M$  with noCAR and niCAR in comparison to 3-phenylpropionic acid. However, in all cases, the  $k_{cat}$  was reduced when compared to 3-phenylpropionic acid or benzoic acid, as it was for (*E*)-3-phenylprop-2-enoic acid. Finally, *trans*-2-phenylcyclopropane-1-carboxylic acid (compound **13**, Figure 4-3C, Table 4-3) features a cyclopropane ring between the benzene ring and carboxylate group. For all the CARs this modification resulted in much lower  $K_M$  values, and a much lower  $k_{cat}$ , compared to 3-phenylpropionic acid or benzoic acid.

#### 4.5.6 Heterocycles

Heterocycles containing nitrogen, oxygen or sulfur were tested (compounds **14-17**, Figure 4-3D, Table 4-4). Generally, weak activity was observed, with decreasing  $K_M$  values for increasing heteroatom size. Where there was activity,  $k_{cat}$  was generally lower than benzoic acid. Possibly, in cases where no activity was detected, the  $K_M$  was outside the range of detection of the assay.

#### 4.5.7 Fatty acids

All the CARs showed very high catalytic efficiency for fatty acids between eight and twelve carbons in length, with low  $K_M$  values compared to benzoic acid (compound **18-21**, Figure 4-3E, Table 4-4). Octadecanoic acid (**21**), with a carbon chain length of 18 carbons, showed a similarly low  $K_M$  but a greatly reduced  $k_{cat}$ . All CARs except mpCAR were active against butanoic acid (compound **18**, Figure 4-3E, Table 4-4) but with a very large  $K_M$ , in most cases too large to characterize accurately. In general, mpCAR was much less efficient with fatty acids than the other CARs.



**Figure 4-3 – CAR activity for various benzoic acid derivatives, heterocycles and fatty acids.**

The  $k_{cat}$  ( $\text{min}^{-1}$ ) determined for each enzyme against each substrate displayed as follows: mpCAR ■, msCAR ■, tpCAR ■, noCAR ■ and niCAR ■. Below each substrate is its chemical structure. Error bars show the standard error. A: Benzoic acid and derivatives with electron donating groups. B: Derivatives with an electron withdrawing groups. C: Derivatives with various substituents between the carboxylate group and benzene ring. D: Heterocycles containing either an oxygen, sulfur or nitrogen. E: Fatty acids between four and eighteen carbons in length.

**Table 4-2 – CAR activity against benzoic acid and its derivatives with electron donating and withdrawing groups**

		1. Benzoic acid	2. 4-methyl-benzoic acid	3. 4-Methoxy-benzoic acid	4. 3-Methoxy-benzoic acid	5. 2-Methoxy-benzoic acid	6. 4-Nitro-benzoic acid	7. 3-Nitro-benzoic acid	8. 2-Nitro-benzoic acid
	<i>Hammet sigma constants:</i> <sup>[26]</sup>	0	-0.17	-0.27	0.12	-	0.71	0.78	-
<b>mpCAR</b>	$k_{cat}$ (min <sup>-1</sup> )	140 ± 20*	122 ± 3	132 ± 4	104 ± 3	NA	NA	3.7 ± 0.5	NA
	$K_M$ (mM)	20 ± 4*	3.7 ± 0.2	2.8 ± 0.2	3.0 ± 0.2	NA	NA	0.3 ± 0.1	NA
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	7 ± 1*	33 ± 2	48 ± 5	35 ± 3	NA	NA	11 ± 4	NA
<b>msCAR</b>	$k_{cat}$ (min <sup>-1</sup> )	197 ± 4	154 ± 6	179 ± 6	18 ± 1*	NA	NA	40 ± 10	NA
	$K_M$ (mM)	3.4 ± 0.2	0.16 ± 0.02	0.19 ± 0.02	12 ± 1*	NA	NA	0.5 ± 0.2	NA
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	57 ± 4	900 ± 100	930 ± 80	1.4 ± 0.2*	NA	NA	100 ± 50	NA
<b>tpCAR</b>	$k_{cat}$ (min <sup>-1</sup> )	142 ± 3	152 ± 2	130 ± 2	186 ± 2	19 ± 3	13 ± 1	33 ± 2	NA
	$K_M$ (mM)	2.0 ± 0.1	0.69 ± 0.03	0.45 ± 0.02	0.56 ± 0.02	9 ± 3	0.6 ± 0.1	0.7 ± 0.1	NA
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	72 ± 6	220 ± 10	290 ± 10	334 ± 10	2.2 ± 0.7	22 ± 6	44 ± 8	NA
<b>noCAR</b>	$k_{cat}$ (min <sup>-1</sup> )	183 ± 6	135 ± 5	138 ± 4	136 ± 5	NA	NA	59 ± 2	NA
	$K_M$ (mM)	2.1 ± 0.2	1.2 ± 0.2	1.1 ± 0.1	0.9 ± 0.1	NA	NA	2.5 ± 0.3	NA
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	89.1 ± 8	110 ± 20	130 ± 10	150 ± 10	NA	NA	24 ± 3	NA
<b>niCAR</b>	$k_{cat}$ (min <sup>-1</sup> )	98 ± 7	94 ± 2	49 ± 1	93 ± 1	NA	NA	18 ± 1	NA
	$K_M$ (mM)	0.9 ± 0.1	1.0 ± 0.1	0.25 ± 0.01	0.68 ± 0.03	NA	NA	5.6 ± 0.7	NA
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	103 ± 9	97 ± 6	200 ± 10	137 ± 6	NA	NA	3.2 ± 0.4	NA
NA: no activity was detected with that substrate. *: $K_M$ was unusually large and substrates concentrations could not reach a high enough concentration to accurately determine kinetic constants. Errors represent the standard error. No Hammett constants for ortho positions as steric effects cannot be properly accounted for.									



Table 4-3 – CAR activity against benzoic acid derivatives with the carboxylic acid group extended from the ring.						
		9. 3-Phenylpropionic acid	10. ( <i>E</i> )-3-phenylprop-2-enoic acid	11. Phenylpropynoic acid	12. 3-oxo-3-phenylpropanoic acid	13. <i>trans</i> -2-Phenylcyclopropane-1-carboxylic acid
mpCAR	$k_{cat}$ (min <sup>-1</sup> )	21.5 ± 0.7	67 ± 2	NA	18 ± 2	20 ± 1
	$K_M$ (mM)	3.0 ± 0.3	0.3 ± 0.02	NA	3.8 ± 0.8	1.8 ± 0.2
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	7.2 ± 0.7	240 ± 2	NA	5 ± 1	12 ± 1
msCAR	$k_{cat}$ (min <sup>-1</sup> )	184 ± 9	118 ± 2	NA	75 ± 2	2.2 ± 0.1
	$K_M$ (mM)	0.16 ± 0.02	0.075 ± 0.004	NA	0.27 ± 0.02	0.006 ± 0.0001
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	1200 ± 200	1600 ± 500	NA	280 ± 20	380 ± 20
tpCAR	$k_{cat}$ (min <sup>-1</sup> )	158 ± 2	38 ± 1	6 ± 4	85 ± 2	43 ± 1
	$K_M$ (mM)	0.32 ± 0.01	0.310 ± 0.002	0.09 ± 0.02	0.55 ± 0.04	0.061 ± 0.005
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	500 ± 20	120 ± 2	70 ± 40	150 ± 10	700 ± 60
noCAR	$k_{cat}$ (min <sup>-1</sup> )	140 ± 4	105 ± 3	NA	63 ± 2	48 ± 1
	$K_M$ (mM)	2.7 ± 0.2	0.72 ± 0.07	NA	0.29 ± 0.03	1 ± 0.1
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	52 ± 4	147 ± 15	NA	210 ± 20	46 ± 3
niCAR	$k_{cat}$ (min <sup>-1</sup> )	85.8 ± 0.9	7.7 ± 0.7	7 ± 0.4	37.1 ± 0.5	10.8 ± 0.3
	$K_M$ (mM)	0.97 ± 0.03	0.05 ± 0.02	1.3 ± 0.2	0.39 ± 0.02	0.21 ± 0.02
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	88 ± 3	170 ± 70	5 ± 1	94 ± 4	51 ± 5
NA: no activity was detected with that substrate. Errors represent the standard error.						

**Table 4-4** – CAR activity against heterocycles and fatty acids

		14. Pyridine-2-carboxylic acid	15. 1 <i>H</i> -Pyrrole-2-carboxylic acid	16. Furan-2-carboxylic acid	17. Thiophene-2-carboxylic acid	18. C4 – Butanoic acid	19. C8 – Octanoic acid	20. C12 – Dodecanoic acid	21. C18 – Octadecanoic acid
mpCAR	$k_{cat}$ (min <sup>-1</sup> )	NA	NA	NA	50 ± 20*	NA	58 ± 1	55 ± 2	3.7 ± 0.3
	$K_M$ (mM)	NA	NA	NA	50 ± 20*	NA	2.0 ± 0.1	0.09 ± 0.01	0.09 ± 0.02
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	NA	NA	NA	1.1 ± 0.6*	NA	29 ± 2	600 ± 70	39 ± 9
msCAR	$k_{cat}$ (min <sup>-1</sup> )	NA	NA	50 ± 10	123 ± 4	129 ± 7*	296 ± 8	131 ± 5	46 ± 4
	$K_M$ (mM)	NA	NA	13 ± 4	3.3 ± 0.3	7.9 ± 0.8*	0.1 ± 0.01	0.05 ± 0.01	0.6 ± 0.09
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	NA	NA	4 ± 2	37 ± 3	16 ± 2*	3000 ± 300	2700 ± 400	80 ± 10
tpCAR	$k_{cat}$ (min <sup>-1</sup> )	23 ± 3	NA	19 ± 1	82 ± 3	82 ± 3*	219 ± 3	157 ± 5	15 ± 1
	$K_M$ (mM)	24 ± 7	NA	4.7 ± 0.5	3.3 ± 0.3	5.0 ± 0.4*	0.2 ± 0.01	0.04 ± 0.01	0.12 ± 0.03
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	0.9 ± 0.3	NA	4.0 ± 0.4	25 ± 3	17 ± 2*	1140 ± 50	3600 ± 400	120 ± 30
noCAR	$k_{cat}$ (min <sup>-1</sup> )	76 ± 4	NA	NA	135 ± 4	170 ± 20*	141 ± 2	99 ± 3	11 ± 1
	$K_M$ (mM)	20 ± 2	NA	NA	2.6 ± 0.2	50 ± 8*	0.2 ± 0.01	0.04 ± 0.01	0.02 ± 0.01
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	3.9 ± 0.4	NA	NA	52 ± 4	3.4 ± 0.7*	750 ± 30	2500 ± 300	500 ± 300
niCAR	$k_{cat}$ (min <sup>-1</sup> )	NA	NA	NA	60.8 ± 0.9	260 ± 30*	233 ± 5	157 ± 9	68 ± 7
	$K_M$ (mM)	NA	NA	NA	1.00 ± 0.05	32 ± 4*	0.2 ± 0.01	0.02 ± 0.01	0.7 ± 0.1
	$k_{cat} / K_M$ (min <sup>-1</sup> mM <sup>-1</sup> )	NA	NA	NA	58 ± 3	8 ± 1*	1350 ± 90	7000 ± 2000	100 ± 20
NA: no activity was detected with that substrate. *: $K_M$ was unusually large and substrates concentrations could not reach a high enough concentration to accurately determine kinetic constants. Errors represent the standard error.									

#### 4.5.8 Effects of pH

The activity of an enzyme at different pH values is an important consideration for an industrial enzyme. Therefore, the effect of pH on CAR activity was examined by measuring activity against benzoic acid at different pH values. mpCAR, niCAR, noCAR and tpCAR all showed optimum activity at pH 7.5, whilst msCAR showed an optimum activity at pH 7.8 (Figure 4-4). Both niCAR and tpCAR show a sharp peak of activity around pH 7.5, with activity quickly decreasing as the pH moved away from this point. In contrast, mpCAR and noCAR show a slightly broader optimum around pH 7.0 to 7.6. msCAR behaves very differently from the other CARs. At more acidic pH values between pH 5.5 and 6.8 it shows very low activity where the other CARs are more active. However, it is also more active at more alkaline pH values where the other CARs are less active.

#### 4.5.9 Effects of temperature

Thermostability was investigated by incubating the CAR enzymes at various temperatures for half an hour and measuring residual activity against 4-methylbenzoic acid relative to a control kept on ice. tpCAR was the least thermostable CAR tested, being completely inactive after half an hour at 42 °C (Figure 4-5A). In contrast, mpCAR, a CAR from the moderate thermophile *M. phlei*, retains 92% of its activity following the same incubation at 42 °C. mpCAR was able to retain residual activity up to 50 °C making it the most thermostable CAR identified to date. Both niCAR and noCAR showed intermediate thermostability, denaturing at temperatures beyond 44 °C, while msCAR is marginally more thermostable and is able to retain some activity until 47 °C.

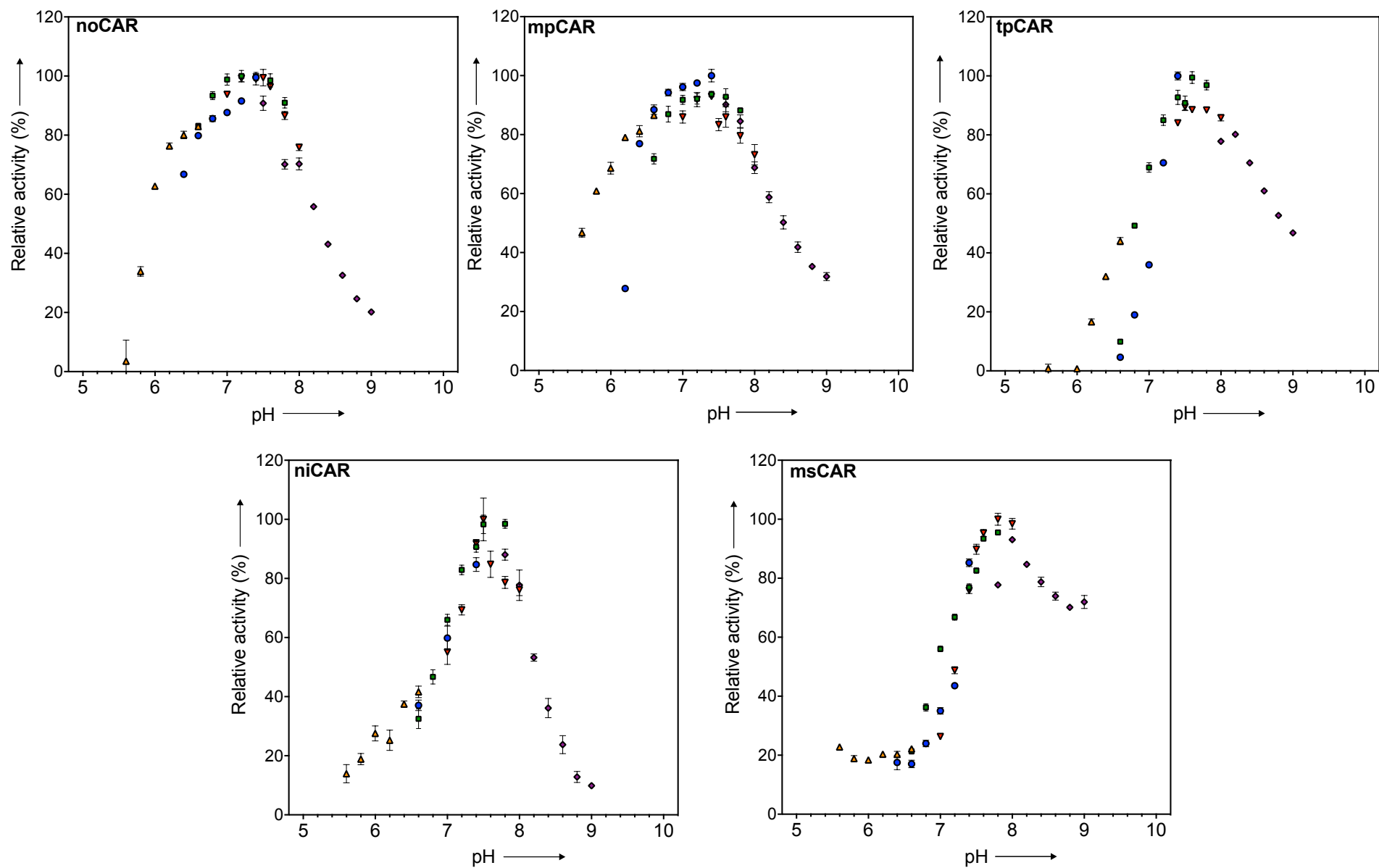
Activity at temperature was tested in a 10 minute reaction. The more thermostable CARs, mpCAR, msCAR and niCAR all showed an optimum activity of 42 °C (Figure 4-5B). Activity decreased past this temperature at various degrees relative to the thermostability of each enzyme. noCAR showed a slightly lower optimum at 38 °C while tpCAR had a much lower optimum still, at only 31 °C.

The half-life and degradation constant at 30 °C were calculated by measuring activity at various time points over 120 hours. The data were fitted to a one phase decay equation by non-linear least squares regression. mpCAR, a CAR from a moderate

thermophile, showed by far the longest half-life at 30 °C at 123.2 hours (Table 4-5). In contrast, tpCAR has a much shorter half-life of only 25.0 hours. The half-lives of msCAR, niCAR and noCAR fell between these extremes at 53.7, 42.9 and 35.3 hours respectively. Total turnover numbers (TTN) for the three best substrates were calculated as  $k_{cat} / K_D$  (Table 4-5).

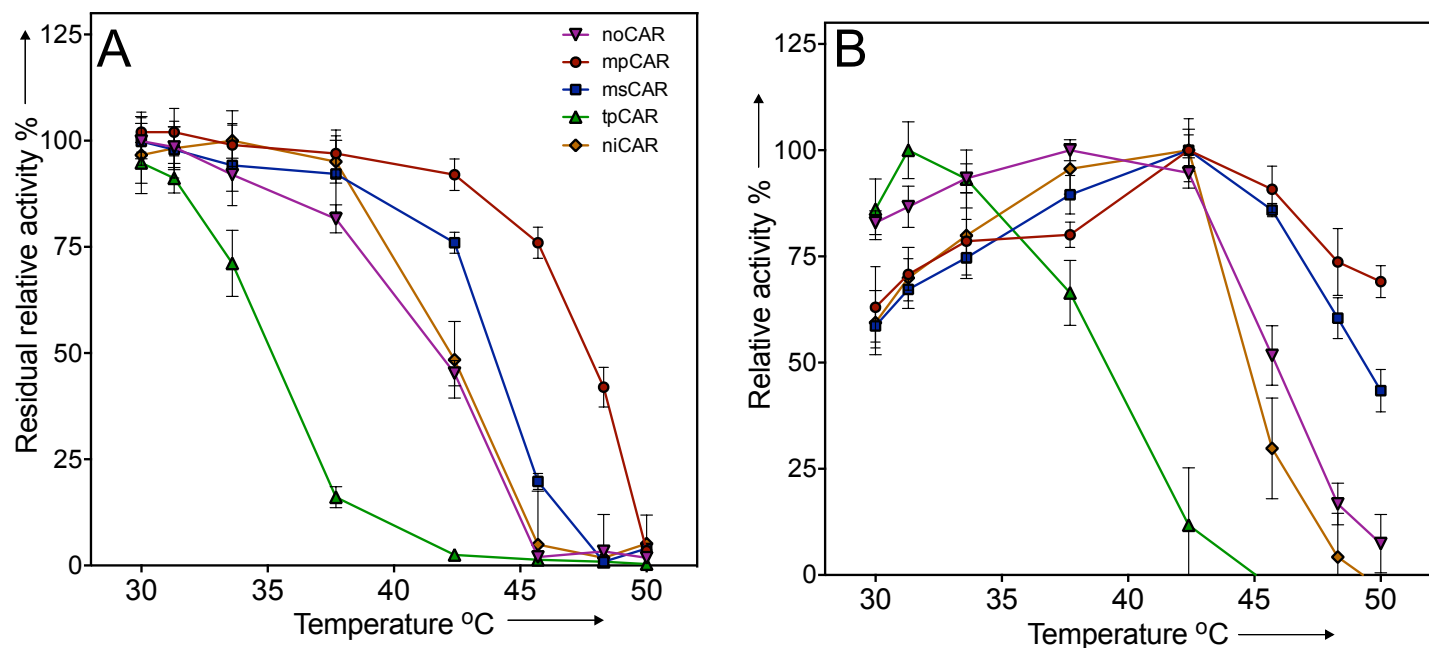
#### **4.5.10 Product inhibition**

mpCAR was tested for product inhibition with AMP, NADP<sup>+</sup> and PP<sub>i</sub>. NADP<sup>+</sup> showed competitive inhibition with NADPH with a  $K_i$  of  $143 \pm 8 \mu\text{M}$  (Supplementary Figure 4-13) and AMP was a competitive inhibitor of ATP with a  $K_i$  of  $8200 \pm 900 \mu\text{M}$  (Supplementary Figure 4-14). PP<sub>i</sub> showed mixed inhibition with ATP with a  $K_i$  of  $220 \pm 50 \mu\text{M}$ , and an  $\alpha$  of  $2.5 \pm 1.4$ . Surprisingly, PP<sub>i</sub> also showed competitive inhibition with 4-methylbenzoic acid with a  $K_i$  of  $340 \pm 40 \mu\text{M}$  (Supplementary Figure 4-15 and Supplementary Figure 4-16).



**Figure 4-4 – The activity of CAR enzymes in response to pH.**

Overlapping buffers were used to cover a range from pH 5.6 to pH 9.0 in intervals of 0.2 and are displayed as follows: MES-NaOH ▲ PIPES-NaOH ●, MOPS-NaOH ■, HEPES-NaOH ▼, Tris-HCl ◆ Activity against 4-methylbenzoic acid is shown relative to the highest activity at 100%. Errors bars show the combined standard deviation of three readings and three blank readings (with no enzyme) at each pH values.



**Figure 4-5 – The effects of temperature on CAR enzymes.**

**A** – Thermostability of CAR enzymes. The residual activity of CAR enzymes against 4-methylbenzoic acid after a 30-minute incubation at different temperatures is displayed. Activity is shown relative to a control sample kept at 4 °C with errors bars showing the standard deviation of three readings. **B** – Activity of CAR enzymes at different temperatures. Activity is relative to the fastest rate at 100%. Error bars show the combined standard deviation of three readings and three blank reading (with no substrate) at each temperature.

**Table 4-5 – Half-life and degradation constant  $K_D$  of CAR enzymes when incubated at 30 °C.**

<b>Enzyme</b>	<b>Halflife (hours)</b>	<b><math>K_D</math> (hrs<sup>-1</sup>)</b>	<b>TTN benzoic acid</b>	<b>TTN 4- methylbenzoic acid</b>	<b>TTN 4- methoxybenzoic acid</b>
mpCAR	123.2	$0.0056 \pm 0.004$	$30000 \pm 20000$	$20000 \pm 20000$	$20000 \pm 20000$
msCAR	53.7	$0.013 \pm 0.001$	$15000 \pm 1000$	$10000 \pm 1000$	$14000 \pm 1000$
niCAR	25	$0.28 \pm 0.002$	$350 \pm 30$	$336 \pm 8$	$175 \pm 4$
noCAR	35.3	$0.02 \pm 0.002$	$9000 \pm 1000$	$6800 \pm 700$	$6900 \pm 700$
tpCAR	42.9	$0.016 \pm 0.002$	$9000 \pm 1000$	$6000 \pm 1000$	$8000 \pm 1000$
The half-life, $K_D$ and TTN for the three best substrates of CAR enzymes calculated from activity after incubation at 30 °C over time, fitted to $Y=Y_0 \cdot e^{-K \cdot X}$ . Standard error for $K_D$ is shown. TTN has been calculated as $k_{cat} / K_D$ , with the combined error shown.					



## 4.6 Discussion

The CAR enzymes offer an excellent opportunity for green chemistry: they offer the opportunity to reduce carboxylic acids selectively to aldehydes, without the use of harsh reducing agents. CARs also have the clear advantage over other enzymes capable of carrying out this reaction of the reduced product being thermodynamically favored, due to the hydrolysis of ATP. Although previous studies have identified a few CARs from different species, and demonstrated that they have activity against diverse acids, none of these studies has provided a detailed, kinetic comparison of diverse CARs. We therefore aimed to thoroughly characterize example CARs from across the known CAR family, together with the best-characterized CAR from *N. iowensis*. Our aim was to demonstrate the similarities and differences between these CARs, learn more about the CAR mechanism, and highlight the potential of these enzymes for biocatalysis.

### 4.6.1 The effect of the addition of electron donating or withdrawing groups

The reduction of carboxylic acids to aldehydes typically involves a transfer of a 'hydride' to the carbonyl unit. We therefore initially expected that electron withdrawing groups, which make this carbon more electrophilic, would be preferred substrates. However, our observation was that, contrary to our expectation, electron donating groups were preferred substrates (Figure 4-3, Table 4-2). The addition of electron donating groups to benzoic acid resulted in a reduction in  $K_M$ , and so an increase in catalytic efficiency. We reasoned that these groups would drive electrons into the  $\pi$ -system, making the first step of the reaction (attack by the negatively charged carboxylate group on the  $\alpha$  phosphate of ATP) more favorable. As two of the other steps (2 and 4) involve nucleophilic attacks on the acid group carbon atom of the carboxyl group (which should favor electron withdrawing groups), this strongly suggests that the first step in the reaction has the greatest impact on substrate specificity and selection. It is possible that the reduced  $K_M$  with electron donating substituents is a consequence of the acyl-AMP intermediate forming more readily, although very detailed studies of the kinetics of this individual step would be required to confirm this. In long-chain fatty acid ligases, the acyl-AMP intermediate has been shown to be unable to leave the active site (Hisanaga *et al.*, 2004), so the addition of

a group which likely improves the formation of this intermediate might be expected to cause a lower  $K_M$  and greater catalytic efficiency. Moreover, when benzyl-AMP was used as a substrate with a CAR from *Nocardia asteroides* it showed a  $K_M$  of 70 nM, compared to 260 nM for benzoic acid, suggesting that this intermediate binds more tightly to the enzyme (Kato *et al.*, 1991). Furthermore, the phosphopantetheine binding and C-terminal reductase domains shows high sequence identity to that of other ANL superfamily members that process very different substrates. For example a NRPS from *Mycobacterium intracellulare*, WP\_014382786.1, has an average of 58 % identity to the CARs in Figure 4-2 for this C-terminal region). This strongly suggests that substrate specificity must be determined in the adenylation domain, likely at the formation of the first intermediate.

In the 3-position, the methoxy group has no resonance effect on the carboxylic acid and so is actually slightly electron withdrawing by induction, as indicated by the Hammett sigma constants in Table 4-2. In many of the CARs, the  $k_{cat}$  of 3-substituted benzoic acids shows a small reduction compared to 4-substituted acids, with msCAR showing greatly reduced activity. However, these are still good substrate for most of the CARs. It is likely that there are further interactions between the substrate and the active site binding pocket, and that electronic effects alone cannot account for all differences in activity.

Very low or no activity was found with a 2-methoxy substituent of benzoic acid. This suggests that there is a steric interference by the methoxy group on the binding of the nearby carboxylate group to the relevant area of the active site. This effect has been reported for other CARs examined to date with other 2-substituents. However, some cases suggest there is activity, but at a low level (Kato *et al.*, 1991; Li and Rosazza, 1997). No structure of a CAR enzyme has yet been described, and this would be highly beneficial in understanding the effects of groups in the 2-position.

All substrates with an electron withdrawing group showed much lower  $k_{cat}$  values than benzoic acid, in most cases inhibiting activity all together. These groups should increase the propensity of the carbonyl carbon to nucleophilic attack in steps 2 and 4 of the reaction. Therefore, this strongly suggests again that these two steps are of limited relevance for substrate specificity. Only 3-nitrobenzoic acid showed activity with all the CAR enzymes, likely as in this position the electron withdrawing group

has no resonance effect on the carboxylate group. As is the case with the methoxy group, it is possible that a 2-nitro substituent inhibits activity due to a steric hindrance because of its close proximity to the carboxylate group.

Previously it has been reported that 2-substituted benzoic acids are poor substrates for niCAR, in good agreement with our data (Li and Rosazza, 1997). However very low activity was observed with 2-methoxybenzoic acid previously, which we did not detect. Substrates with the addition of electron donating groups to benzoic acid were previously shown to be good substrates for niCAR, in agreement with our results. The activity of niCAR with electron withdrawing chloro and bromo 3-substituted benzoic acids supports our reasoning that in the meta position the absence of a resonance effect allows better activity with these substrates than in the other positions (Li and Rosazza, 1997).

#### **4.6.2 Modifications between the benzene ring and the carboxylic acid group.**

3-Phenylpropionic acid disrupts the influence of the aryl unit on the carboxylic acid compared to benzoic acid, and also makes a less sterically rigid substrate. This difference seems to have made the carboxylic acid group more accessible, as the  $K_M$  is much lower than benzoic acid in most cases. The inclusion of a double bond in (*E*)-3-phenylprop-2-enoic acid, should withdraw electrons from the carboxylic acid group. This would be beneficial for nucleophilic attack on the carbonyl in steps 2 and 4 of the reaction, but detrimental to the initial attack by the oxygen of the carboxylate group on ATP. The inclusion of the double bond causes a significant drop in  $k_{cat}$ . The double bond also makes the molecule more rigid, in an apparently favorable conformation, as the  $K_M$  is even smaller than for 3-phenylpropionic acid.

When a triple bond is added to the structure (phenylpropynoic acid), the molecule is very rigid and flat, with a more electron deficient carboxylic acid group. These effects together removed activity in nearly all the CARs. The presence of a  $\beta$ -ketone group into the  $\beta$ -carbon of 3-phenylpropionic acid will have a similar effect to the inclusion of a double bond in (*E*)-3-phenylprop-2-enoic acid, very weakly withdrawing electrons from the carboxylic acid group. Indeed, the  $k_{cat}$  values for phenylpropynoic acid and (*E*)-3-phenylprop-2-enoic acid are similar. The ketone group had mixed

effects on the  $K_M$  for the various CARs, suggesting differing interactions taking place with the ketone group within the active sites of the enzymes. These observations all agree with the hypothesis that the first step of the proposed reaction mechanism is rate limiting.

### 4.6.3 Heterocycles

The CARs generally showed less activity towards heterocycles as towards a benzene ring. They showed a preference for heterocycles containing a larger heteroatom, or with a less aromatic nature. In Thiophene-2- carboxylic acid, the lone pairs of electrons in the sulfur atom are more dispersed and less available for bonding, which possibly results in the lower  $K_M$ . In contrast, the nitrogen atom in 1H-pyrrole-2-carboxylic or pyridine-2-carboxylic acid has lone pairs more available for bonding, which may result in the very large  $K_M$  values, or lack of activity observed. Furan-2-carboxylic acid, with an oxygen atom in the heterocycle, sits between these substrates in both respects.

### 4.6.4 Fatty acids

Fatty acids make interesting substrates, since fatty alcohols can be used as biofuels, in detergents, surfactants and polymers (Napora-Wijata, Strohmeier and Winkler, 2014). As was observed for the CAR from *Mycobacterium marinum* (mmCAR), most of the CARs tested were active against fatty acids between C4 and C18, with similar kinetics to previous work observed (Akhtar, Turner and Jones, 2013). Catalytic efficiency with butyric acid was very poor, primarily due to large  $K_M$ 's for this substrate suggesting it might be too small to make the necessary interactions in the active site of the adenylation domain. However, larger fatty acids showed much lower  $K_M$ 's, with high turnover numbers resulting in catalytic efficiencies higher than any of the aromatic substrates tested in many cases. As the acyl chain length increased past octanoic acid (with an eight carbon chain length),  $k_{cat}$  decreased, reaching a low residual level for stearic acid. Both niCAR and msCAR showed a better turnover number with the longer steric acid than the other CARs, suggesting these enzymes might be better suited to larger substrates. Recently two other CARs, in combination with niCAR and mmCAR have been shown to have activity against ethanoic, butanoic, 2-methyl butanoic and 2-oxobutanoic acids, highlighting

that CARs can accept small fatty acids and that they can tolerate the addition of groups such as a methyl or carbonyl group onto the alpha carbon. However, 2-aminobutanoic acid was also tested but showed no activity (Moura *et al.*, 2015).

#### **4.6.5 Effects of pH and temperature**

The operating pH and temperature range of an enzyme is an important consideration for a potential biocatalyst. Stability at extremes of pH and in solvents are characteristics often found in thermostable proteins, as the mechanisms stabilizing these proteins against high temperature can also be stabilizing against these other conditions. We observed an optimum pH of 7.5 for four of the five CARs tested, with a general tolerance to acidic pH, consistent with previously reported data on the activity of other CARs (Kato *et al.*, 1991). In particular, both mpCAR and noCAR were able to tolerate pH 6 with only a small loss of activity (whilst other CARs showed a much narrower optimum). In contrast to this, msCAR is clearly better suited to more alkaline pH values (Figure 4-4). This therefore offers a CAR suitable for use in biocatalysis in conjunction with other enzymes favoring a similarly alkaline pH.

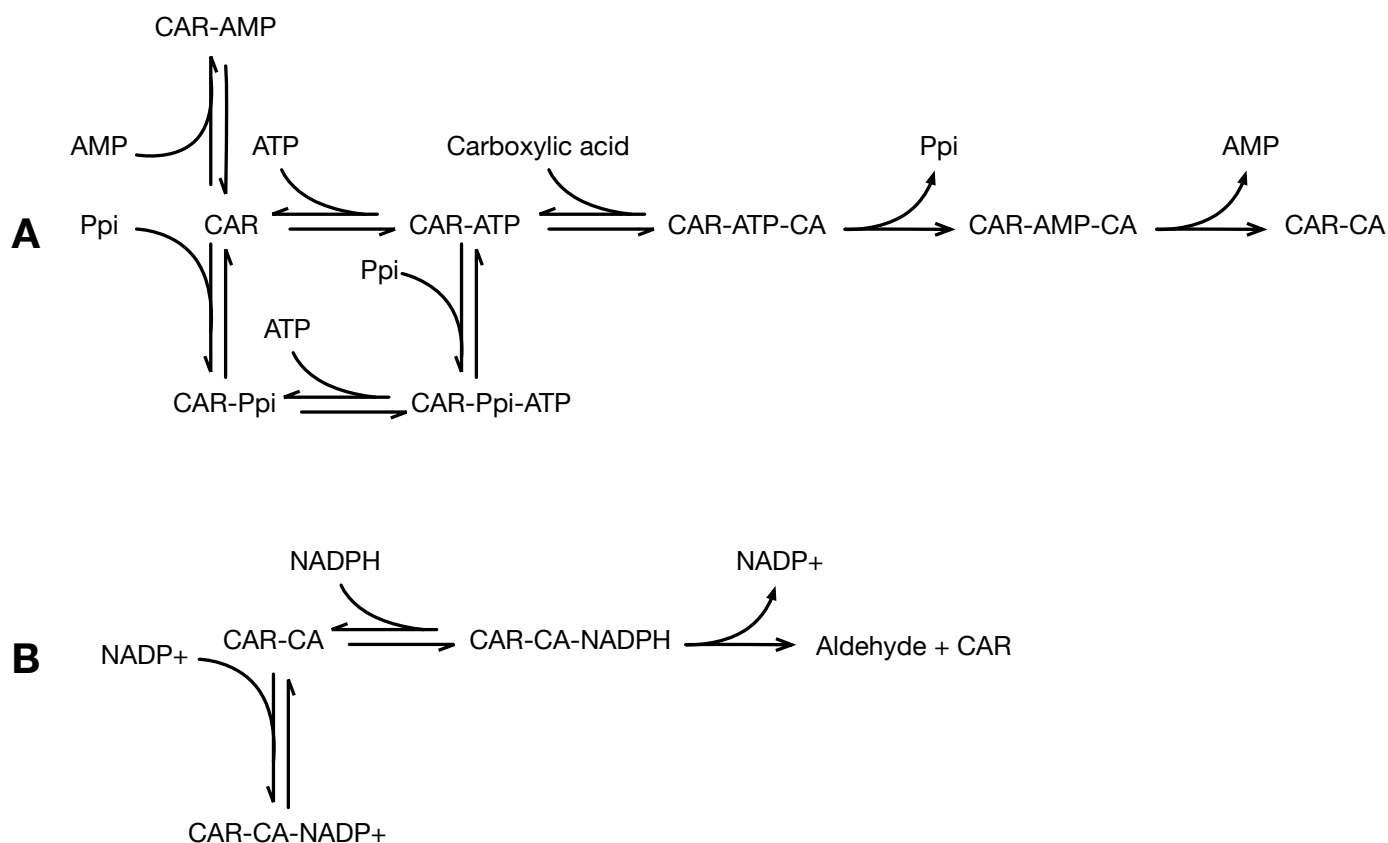
mpCAR showed by far the best thermostability of any characterized CAR (Figure 5, Table 6). We also observed that it shows a much lower catalytic efficiency in general than the other CARs at 30 °C (Tables 2-4). Possibly, there has been a trade-off between the rigidity of the enzyme (providing thermostability) and flexibility to allow a broader substrate range. It was notable that the rate enhancement in mpCAR at its optimum temperature compared to 30 °C was little greater than that for other CARs (Figure 4-5B). In contrast, tpCAR shows very poor thermostability (Figure 4-5A), but is active with many of the substrates that the others CARs could not turn over (e.g. compounds 5, 6, 11, 14). A possible compromise enzyme is msCAR, which shows the next best thermostability, and also has generally good catalytic efficiency. When choosing an enzyme for industrial use, the lifespan of the enzyme can be an important consideration. TTN can be calculated as a measure of how effective an enzyme will be over its lifetime, which we have demonstrated with three of the best CAR substrates (Table 4-5). In this respect the most thermostable CARs have an obvious advantage in that the total turnover number of these enzymes will be much greater (Rogers and Bommarius, 2011). We observed that the lifespan of the

enzyme at 30 °C (Table 4-5) mirrored the thermostability of the enzymes exactly (Figure 4-5), suggesting that a test of thermostability will be a good predictor of lifespan for CARs.

#### **4.6.6 Product inhibition and reaction mechanism**

mpCAR was shown to be inhibited by most of its reaction products and it is assumed that the other CARs share this inhibition. It is unsurprising that  $\text{NADP}^+$  acts as a competitive inhibitor of NADPH (Supplementary Figure 4-13) as  $\text{NADP}^+$  is likely also able to bind to the Rossmann fold of the reductase domain. AMP acts as a competitive inhibitor against ATP (Supplementary Figure 4-14), likely as they are very similar molecules. AMP has also been shown to be a competitive inhibitor of ATP in long-chain fatty acid CoA synthetases, in which the adenylation domain shows significant homology to the CAR adenylation domain (Li *et al.*, 2007).

$\text{PP}_i$  showed mixed inhibition against ATP but competitive inhibition against 4-methylbenzoic acid (Supplementary Figure 4-15 and Supplementary Figure 4-16). This pattern of inhibition is characteristic for ordered sequential bisubstrate reactions (Cook and Cleland, 2007). This indicates that ATP is first to bind to the adenylation domain and is then followed by a carboxylic acid. Long-chain fatty acid CoA synthetases show the same ordered binding of these substrates (Hisanaga *et al.*, 2004). We therefore propose a model for the ordered binding of substrates and inhibitors to the CAR enzyme based on these results (Figure 4-6). It is also interesting that whilst  $\text{PP}_i$  is a product of ATP, its activity as an inhibitor shows that it preferentially binds to the carboxylic acid binding site. CARs might therefore need to be combined with other enzymes such as phosphite dehydrogenase (Zou *et al.*, 2012) or inorganic pyrophosphatase (Lahti, 1983) to overcome product inhibition in an industrial process.



**Figure 4-6 – Model for binding of substrates and inhibitors to the CAR enzyme.**

**A:** binding and release of substrates, products and inhibitors in the adenylation domain. The final result is the formation of a thioester intermediate with the phosphopantetheine arm, represented by CAR-CA. The phosphopantetheine arm can then transfer CA to the reduction domain, **B**, where it is reduced by NADPH, releasing the aldehyde product.

#### 4.6.7 CAR phylogeny

Here we have provided the first glimpse of CAR evolution within the *Actinomycetes*. From the phylogeny it can be hypothesized that the CARs may have propagated through the *Nocardia* and *Mycobacteria* by a series of early horizontal transfer events. This is most apparent in *M. smegmatis*, which possesses three CAR paralogues that cluster in two distinct Mycobacterial clades. Additionally, it is apparent that a large amount of change has occurred within the *Tsukamurella*. This could reflect the slightly more promiscuous substrate range of tpCAR.

#### 4.6.8 Insight into the adenylation step from the ANL superfamily of enzymes.

We have presented evidence that the adenylation domain of the CARs belongs to the ANL superfamily of enzymes due to the presence of conserved hallmark motifs. This placement is further supported by nonribosomal peptide synthetases (NRPSs) and the acyl-CoA synthetases both using an acyl group to form a thioester between a substrate and a pantetheine thiol. Furthermore, the NRPSs mobilize their substrate following thiolation of a phosphopantetheine arm that is bound to a *holo*-acyl carrier protein domain. Parallels can be drawn between both above reactions and the proposed mechanism of CAR activity in Figure 4-1. Due to the lack of crystal structure, previous evidence for the functional dynamics of carboxylic acid reduction in CARs was based on its substrate properties and overall sequence structure (Napora-Wijata, Strohmeier and Winkler, 2014). It follows that knowledge of the extensively studied ANL superfamily can be used to offer insight into the finer details of the mechanism of carboxylic acid reduction employed within the CAR family.

ANL superfamily members, which the CAR adenylation domain shows homology to, are further partitioned into two subdomains – a large (~450 aa) N-terminal domain and a small (~100 aa) C-terminal domain, connected by a flexible linker. Crystal structures show the substrate binding pocket is formed by the N- and C-domain interface. Substrate adenylation progresses in a two-step manner, where following the formation of an acyl-bound intermediate and the release of PP<sub>i</sub>, the active site undergoes massive conformational changes due to a ~140° rotation of the C-terminal domain. Within the NRPSs and the acyl-CoA synthetases, the second domain architecture facilitates thiolation of the phosphopantetheine. Lysines that are



required within each active site are positioned on opposing faces of the C-terminal domain, and are conserved within the CARs (Supplementary Figure 4-1). (Sundlov *et al.*, 2012) This suggests that the CARs also undergo characteristic ANL superfamily domain-alteration between steps 1 and 2 (Figure 4-1) to catalytically isolate the adenylation and thioester forming reactions (Gulick, 2009).

Within the NRPSs, the  $\sim 140^\circ$  rotation of the adenylation domain C-terminal further serves to move the holo-acyl carrier protein domain such that the phosphopantetheine arm can interact with the adenylation domain active site in its thioester forming conformation (Mitchell *et al.*, 2012). Substrate loading induces a curl in phosphopantetheine arm architecture, forcing the substrate to disassociate from the adenylation domain and be held in a position favorable for a second domain to interact (Goodrich, Harden and Frueh, 2015).

#### 4.6.9 Summary

The place of CARs in the ANL superfamily gives support to the current model of multi-step carboxylic acid reduction, and provide tantalizing details about the potential mechanistic processes that have to be undertaken by the CARs in order to activate their somewhat unreactive substrate. The conserved functional motifs characteristic of the ANL superfamily identified in the CARs suggest that major conformational changes facilitate the formation of the phosphopantetheine-substrate thioester, as well as the movement of this intermediate to the reductase domain.

Previous work has primarily focused on screening for CAR activity with different substrates at only one substrate concentration, with data provided as relative activity or simply a binary determination of activity (Li and Rosazza, 1997; Moura *et al.*, 2015). Our comparative enzyme kinetic analysis across the CAR family allows substrate preference to be examined in greater depth, and allows some predictions to be made about the likely substrates that these enzymes could usefully turnover industrially. Our data indicates that the CARs have similar substrate preferences, with some orthologues being more promiscuous than others. The substrate preferences strongly suggest that the first step in the proposed reaction mechanism, during which an AMP-carboxylic acid phosphoester intermediate is formed with the release of  $PP_i$ , is critical for determining suitable substrates. Consequently, the

addition of groups that donate electrons, making the oxygen of the carboxylic acid more electronegative, will be better substrates. Inhibition studies suggest that this first step is an ordered sequential Bi Bi reaction, with ATP being loaded before the carboxylic acid.

The CAR enzymes in this study show only moderate thermostability. To date no CAR enzymes have been identified in any thermophilic organisms. A thermostable CAR enzyme would be attractive for use industrially as this enzyme would likely be resistant to other denaturing forces such as extremes of pH or solvent and likely offer a higher total turnover number for use *in vitro* reactions.

All of the by-products produced in the reaction ( $\text{PP}_i$ , AMP and  $\text{NADP}^+$ ) appear to be inhibitors, and so for the use of CARs *in vitro* there is a need to remove or regenerate all these by-products.

For the application of CAR enzymes in a cascade reaction with other enzymes, pH optimum is also an important consideration. Whilst some of the CARs showed activity across a broader range of pH values than others, various CARs are available to cover reactions with good activity between pH 6-9 at least.

Our results have validated CARs as a useful tool for novel biocatalysis and could be integrated with other enzymes for efficient reduction of carboxylic acids to aldehydes *in vitro*.

## 4.7 Methods

### 4.7.1 Alignments and Phylogeny Construction

Unless specified, all algorithms were performed under default settings. 48 sequences were retrieved by homology search in BLAST to the *N. iowensis* CAR. Alignments were performed using the MUSCLE plug-in within Geneious version 9.1 (<http://www.geneious.com>) (Kearse *et al.*, 2012). Sequence masking was conducted with the Gblocks algorithm within the Phylogeny.fr online tool (<http://www.phylogeny.fr>) (Dereeper *et al.*, 2008). ProtTest (version 3.4)(Abascal, Zardoya and Posada, 2005) analysis of the aligned dataset was performed in the command line. MrBayes (version 3.2.6)(Ronquist *et al.*, 2012) was run in the command line as follows: The amino acid substitution model was fixed to WAG with a gamma-distributed rate variation across a proportion of invariable sites and 8 gamma categories. The analysis was run for 1,000,000 MCMCMC generations, sampling every 100 generations with two parallel runs and four chains (containing one heated chain of temperature 0.2), with a burn-in of 25%. Trees were visualised, midpoint rooted and modified in FigTree version 1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

It must be noted that a more complete list of 124 CAR homologues was retrieved (Supplementary Figure 4-17 and Supplementary Figure 4-18), however a reduced set of sequences was used as this allowed the construction of a more reliable phylogeny.

### 4.7.2 Expression and purification

CAR genes (except niCAR) were cloned into expression vectors pNIC28-Bsa4 (Hansch *et al.*, 1977) or obtained from Prozomix, cloned into pET28a (Novagen). A pET plasmid for the expression of niCAR was obtained from Andrew Hill (University of Manchester). All contained a N-terminal 6x histidine tag (Hansch *et al.*, 1977). Vectors were transformed into BL21 (DE3) *E. coli* along with a pCDF-Duet1 vector containing a phosphopantetheine transferase from *Bacillus subtilis* for its co-expression with the CARs. Expression was carried out in LB media with the addition of 50 µg/µl each kanamycin and spectinomycin. Cells were grown to approximately 0.6 OD<sub>600nm</sub> at 37 °C with shaking at 225 rpm, at which point IPTG was added to a

concentration of 150  $\mu$ M and temperature was dropped to 20 °C for protein expression overnight. Cells were harvested by centrifuging and re-suspended in 25 mM Tris-HCl pH 8.0, 0.5 M NaCl. Cell lysate was prepared by sonication on ice followed by centrifugation to remove the insoluble fraction.

CARs were purified from the cell lysate using a 1 ml His-Trap FF crude column (GE Healthcare) using an elution gradient from 10 to 250 mM imidazole in 25 mM Tris-HCl pH 8.0, 0.5 M NaCl. The purified sample was then applied to a Superdex 200 HiLoad 16/60 gel filtration column (GE Healthcare) and eluted in 25 mM HEPES, pH 7.5, 0.1 M NaCl at 1.0 ml/min. Eluted fractions were analyzed by SDS-PAGE before being pooled and concentrated to approximately 2 mg/ml. To calculate protein concentration from OD<sub>280nm</sub>, an extinction coefficient and molecular weight for each enzyme was calculated using the ExPaSy ProtParam tool, and are shown in Supplementary Figure 4-19. Yields of approximately 2 – 10 mg purified protein per liter of culture were obtained, with 2 – 4 L of culture prepared per batch. Single use aliquots of protein were stored at -80 °C.

#### **4.7.3 Standard enzyme assay**

Unless otherwise specified, assays were carried out in 100 mM Tris-HCl pH 7.5 prepared at 30 °C, 1 mM ATP, 0.25 mM NADPH, 10 mM MgCl<sub>2</sub>, 2-6  $\mu$ g of purified CAR enzyme and 5 mM carboxylic acid substrate in a total volume of 200  $\mu$ l.

Carboxylic acid substrates were prepared in DMSO at 500 mM. The oxidation of NADPH was used to monitor the reactions by measuring the absorbance of NADPH at 340 nm. Reactions were performed in triplicate in a 96-well microtitre plate using a Tecan M200 plate reader at 30 °C over the course of 5 or 10 minutes, after a 5 minute preincubation at 30 °C. Where convenient, an EpMotion 7050 (Eppendorf) liquid handling robot was used to set up the assays.

#### **4.7.4 Kinetic analysis of substrate specificity.**

Kinetic analysis was performed by picking eight appropriate substrate concentrations around an approximate  $K_M$  value for each substrate, and measuring initial rates as described previously. Rates were fitted to the Michaelis-Menten equation by non-linear least squares regression using GraphPad Prism v. 5.0. To calculate constants for ATP and NADPH, 5 mM (*E*)-3-phenylprop-2-enoic acid was used as the

carboxylic acid substrate, except for niCAR where 5 mM 4-methylbenzoic acid was used.

#### **4.7.5 pH vs activity.**

Buffers were prepared and titrated to the correct pH using NaOH or HCl whilst at 30 °C, covering pH values in intervals of 0.2. The buffers 50 mM MES pH 5.6 to 6.6, 50 mM PIPES pH 6.4 to 7.4, 50 mM MOPS pH 6.6 to 7.8, 50 mM HEPES pH 7.0 to pH 8.0 and 50 mM Tris pH 7.8 to pH 9.0 were used. Reactions were carried out as standard with 1 mM ATP, 0.25 mM NADPH, 10 mM MgCl<sub>2</sub> 2-6 µg of purified CAR enzyme and 5 mM 4-methylbenzoic acid. Blanks containing no enzyme were used to subtract a blank rate at each pH value. Initial rates were calculated as relative activity against the fastest result at 100 %.

#### **4.7.6 Thermostability**

A solution containing 2 µg of purified enzyme, 0.25 mM NADPH, 1 mM ATP, 10 mM MgCl<sub>2</sub>, 100 mM Tris-HCl pH 7.5 was incubated across the temperature gradient of a Biorad thermocycler from 30 °C to 50 °C for 30 minutes. The sample was cooled and assayed for CAR activity against 4-methylbenzoic acid, in comparison to a control sample that remained on ice.

#### **4.7.7 Degradation at 30 °C**

2 ml samples at 2 mg/ml in 25 mM HEPES, pH 7.5, 0.1 M NaCl were incubated at 30 °C over a 120 hour period. At specified time intervals, samples were taken and assayed for enzyme activity against 4-methylbenzoic acid. Rates were calculated relative to the first reading at 100 %, and fitted to a model of first order thermal deactivation using the equation  $Y = Y_0 * e^{-K*X}$  where Y is the relative activity and X is the time in hours.

#### **4.7.8 Temperature vs activity**

100 mM Tris-HCl pH 7.5 was prepared at assay temperatures between 30 and 50 °C. Assays were performed as for the thermostability experiment using the temperature gradient of a Biorad thermocycler from 30 °C to 50 °C over the course of 10 minutes, before rapidly cooling on ice with the addition of 10 mM NaOH. A

blank reaction with no substrate was used to calculate the NADPH used in the reaction. Activity was calculated relative to the maximum rate at 100 %.

#### **4.7.9 Product inhibition**

Potential inhibitors were titrated across a broad range of concentrations to determine whether inhibition occurred and to give an idea of an approximate  $K_i$ . Kinetic analysis then was performed as described above using substrates that each inhibitor was likely competitive against, with the addition of the inhibitors at a range of concentrations based around the approximate  $K_i$ . Data were fitted using GraphPad Prism 5.0 by non-linear least squares regression to different models of enzyme inhibition. The model with the best fit for the data was used to determine the mode of inhibition. Where inhibition was not competitive, additional analysis was carried out with other substrates

## 4.8 Acknowledgements

The authors thank Andrew Hill (University of Manchester) for providing many of the substrates tested, the pCDF-Sfp plasmid and the plasmid for the expression of niCAR; and Clive Mountain (GSK), Stacy Clark (GSK), and Alison Hill (University of Exeter) for advice on the chemistry of the CAR reaction, and Jennifer Farrar (Georgia Institute of Technology) for providing walltime on her server to run the Bayesian analyses. Nzomics (Prof. Gary Black and team) and Prozomix (Simon Charnock and team) are gratefully acknowledged for cloning msCAR and tpCAR. WF was funded by BBSRC and GlaxoSmithKline; AT was funded by BBSRC.

## 4.9 Supporting Information

### CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Conserved motifs between adenylating domains of Actinomycete CARs, FadD13 and NPRS

p=medium aliphatic amino acid. r=aromatic amino acid. x=any amino acid.

```
Firefly_Luciferase|BAL46512.1
FadD13|NP_217605.1
NRPS|WP_014382786.1
Tsukamurella_paurometabola
Mycobacterium_smegmatis1
Mycobacterium_genavense
Mycobacterium_marinum
Nocardia_asteroides
Nocardia_paucivorans
Mycobacterium_fortuitum
Mycobacterium_smegmatis2
Mycobacterium_vaccae
Mycobacterium_obuense
Nocardia_seriolae
Nocardia_ottitidiscaviarum
Nocardia_iowensis
Nocardia_transvalensis

-----MEDAKNIKKGPAPFYPLEDGTAGEQLHKAMK-----
-----MKNIGWMLRQ
-----SIPALFAA
---MSIET-----VQNGVPAEGSVPPADQQTERLPQVIARIFAQFAD
---MTSDVHDEQSTRRIAELYATDPEFAAAAPLPAVVDAAHKPLRLAEILQTLFTGYGD
---MTSDNRDERTARRVAELFNNDPQFRAAAPLPEVIEAACAPGLRLTEVLARLVEGYAD
MSITCVDTRAQRSARRIEQLYSTDAQFAAARPSTAVGIAISKSGGLPQIIQTVMDBGYPQ
---MTVEVDADRLADRIRALYAQDAQIRAATPIPEAHARVTTTPTPLARIVSTVMTAYAD
---MSVDTRRESRLERRIAELYATDEQFAAARPDVAVTAEVERAGRRSARVVHAVAKGYAH
---MSFDTRDEQLATRIADLTATDTPQFAAAIPSDTVTASVDVPGLLLPEIVQRVLEGYAE
---MTIETREDRFRNRIDHLFETDPQFAAARPDFAISAAAADPELRLPAAVKQILAGYAD
---MSTDTRREGRLARRIADLFATDPQFAAAVPDETVAAAVEEHAHLDPIMRTVLTDGYAD
---MPTDTRERLAHRIDDLSATDTQFAAALPDEAIAEAIEDPQLRLPQIIATVLDGYAD
---MVEDTARAEIHRRIAERVLADEQVRVAMPLPEVSEAAARQPGGLLARAVEVLMVGYAE
---MLDDARAERERRRIADALADD-QVREAAADAAVSESVRRVEVRLARIVDAVMMSGYGD
---MAVDSPPERLQRRIAQLFAEDEQVKAARPLEAVSAVSAAPGMRLAQIAATVMAGYAD
---MEITDAQAQLIRRATELIEGDEQVRAALPDEAVAKAVQAPGLGLASVVATIMEGYAD

p(S/T)rx(E/Q)p N-terminal helix cap (A1)
RYALVPGTIAFTDAH-----IEVNITYAEYFEMSVRLAEAMKRYGLN-TNHRIV
RATVSPRLQAYVEPS-----TDVRMTYAQMNALANRCADVLTALGIA-KGDRVA
QVARGPGAVAITCGG-----RSFTYRHLYEATNRLAHLHVERGAG-PGQRVA
RPAFATREAGPGTP-----YATVSYREIWRVVTALVASWQSE-VA-PGDFVA
RPALGYRARELATDE-GGRTVTRLRLPRFDTLTYAQVWSRVQAVAAALRHNPIY-PGDAVA
RPALGERVRELVTDA-DGRTVLRRLRPRFETISYRDVWDRVRAIATAWSSDPVT-AGDVVA
RPALGQRATRVVTPDPTGRSSAQLLAEFETITYRELWNRTNALTNAFAAEALADRGRVC
RPALGVRRTELVLVEA--GRATRRLLEFELLTYGEVWERARALAASWYAEGLA-AGEFVA
RPALGQRAVDYITDPTGRSMELLPRYETLTREYVWERAGAIASALAGNPVR-AGDRVC
RPALGERALEFVADPATGRTTARLLPRFDTISYGQVWDRVRALAAALHASGVA-AGDRVA
RPALGKRAVEFVTDE-EGRTTAKLLPRFDTITYRQLAGRIQAVTNAWHNHPVN-AGDRVA
RPALARRAVRFVEDA-GRTVLAELLPHFETITYAELAHRIHGVTALTLD--VH-PGDRVA
RPALGQRAVRLVADPHTERTEAQLLPHFDTITYGELSTRIHLLTALTD--VD-PGDRVA
RPAIGERAEIVTGA-DGRRIRRLLEPYRTITYAELWSRAGAIAAAWQHDPLR-AGDFLC
RAALAWRRSELVDGA-----VRLLEPYSTMITYRELWRQAGAVAAEWGADPVR-AEDFVC
RPAAGQRAFELNTDDATGRTSRLRLPRFETITYRELWQRVGEVAAAWHHDPLR-AGDFVA
RPAAGRAVRFVADP-SGRRHARLLPRYDTITYGELWERVRALMAAHHHDPLR-AGDFVA

structural element (R/K/F)pGp poorly conserved with ANL superfamily (A2)
VCSENSLQFFMPVLGALFIGVAVAPANDIYNERELLNSMNISQPTTVFVSKKGLQKILNV
LLMPNSVEFCCLFYGAALKGAVAVPINTRLAAPVSVFILSDSGSKVVIYGAPSAP-----
VAVPRSAEAIVAILAVLKTGAAYVPIDPSVPAARVQFVLGDAPIAAVTTAEV-----
ILGFTSSDFVTVDLATTLLGAPNVPLQAGAPAARIATILDETRPKILAVSADQVDLAQEA
TIGFASPDYLTLDLVCAVYLGILVSVPLQHNAPVSRAPILAEVEPRILTVSAEYLDLAVES
TVGFSSADYLVVDLVCAVYLGILVTVPLQHNAPPARLRPIIECEPKIVAVSAEYLDLAES
VLGFASIDYATIDIALMLLGAHSVPLPTNAARAQLCHIVSETQPSLIASSTENLPDAISL
TLGFTGADYIVTDLATIIHLGAVAVPLQAGASATQLRSILDETAAPRVLAVDNTANLAVALDV
ILGFASVDYTTIDMALTRLAAVSVPLQTGAPAEERLRPIIAETEPTVLAASIDHLLDDAVTA
ILGFTSADYTVIDTALGQIGAVSVPLQTSSSPEALAPIVTETEPVIAASVDHLADAVEL
ILGFTSVDYTTIDIALLELGAVSVPLQTSAPVAQLQPIVAETEPKVIASSVDFLADAVAL
LLGFTSPDYTVVDMALSMGLAVLVPLQTSAPLSTLRPIIAETEPVLIASSVDTLDDAVAL
ILGFTSVDYTVIDTTLVLRGAVSVPLQTSAPAAATLRPIVAETEPVVFVFAASVDHLSDAVDL
VLGFGSGDFAALEIAAIRQGLVTVPLQANAAAAQWRSIIETGARTLAVSLELLDSALDV
TLGFTSPDYTVVDMALMLRLAAVAVPLQASASVAQWRSIMAETEPRLMAASAETLPAAVEA
LLGFTSIDYATLDIADIHLGAVTVPLQASAAVSQLIAILTETSPRLLASTPEHLDAAVEC
ILGFTGIDYTVVDLACAHLGAVSVPLQAGASLAQLTPIAAETEPVRLATDIEQLGAGVDL

Firefly_Luciferase|BAL46512.1
FadD13|NP_217605.1
NRPS|WP_014382786.1
Tsukamurella_paurometabola
Mycobacterium_smegmatis1
Mycobacterium_genavense
Mycobacterium_marinum
Nocardia_asteroides
Nocardia_paucivorans
Mycobacterium_fortuitum
Mycobacterium_smegmatis2
Mycobacterium_vaccae
Mycobacterium_obuense
Nocardia_seriolae
Nocardia_ottitidiscaviarum
Nocardia_iowensis
Nocardia_transvalensis

QKKLPPIIQKIIIMDSKTDYQGFQSMYTFVTSHL--PPGFNEYDFVPESFDRDKTI-----
```



FadD13|NP\_217605.1  
 NRPS|WP\_014382786.1  
 Tsukamurella\_paurometabola  
 Mycobacterium\_smegmatis1  
 Mycobacterium\_genavense  
 Mycobacterium\_marinum  
 Nocardia\_asteroides  
 Nocardia\_paucivorans  
 Mycobacterium\_fortuitum  
 Mycobacterium\_smegmatis2  
 Mycobacterium\_vaccae  
 Mycobacterium\_obuense  
 Nocardia\_seriolae  
 Nocardia\_otitidiscaviarum  
 Nocardia\_iowensis  
 Nocardia\_transvalensis

-----VIDAIRAQADPPGTVDWIGADSLAERLRS-AADE  
 -----RPQLGGF-----AGQIIDIDDDPAVVRQPATGL-PVPS  
 LAESATPRVVVFD-----GERDGYEGIEADILSGSAL-PAPE  
 VRDVNSVSQVLVVDHHPVEDDHRDALARAREQL--AGKGIATVTLDAIDEGAGL-PAEP  
 ALTSTSLRQLMVFDYRAEVDEQRENFEQTRVRLQGSQTRVAVTTVDEVVARGRL-PAVA  
 VLSHRAPHRVVFDYRPELDHREALEAARARL--AIPVTVETLTAI IARGTRVPAEA  
 VLAGAAPRALVVDHADDNDREVLAARARLRAANSPIVLSTVAEVIDRGRAL-DPAF  
 VLAGHLPARLIVFDYHPRVDDQREAFAAATEKLAEGGPVVIETLAEVIDRGTAL-PPVP  
 ALTAHAPAQLVVDHHPVIDDREAVASAAERITAAGASIAVDTLAGLLDRGSNL-PAPE  
 VESGPAPSRVLVVDYSHEVDDQREAFEAAGKGL--AGTVVVTITDALDRGRSL-ADAP  
 ALDAPDAARLVVDHRAEVDDHRDALTSATARLRAAGSPLEIETLAEVIARGSTM-PARE  
 VADAEVGRILVFDYRAEVDDHRDAIADARARLADAGRSIEIVTLSEVLHAGATL-PAAG  
 VLDGSPVTSIVVDFEPEEDRQAEILVGARDRIAASGSTITLESAAVLERGATL-PAVP  
 VLGGFAPRRVLVFDYRPELEAHRSAVDSARERLAEVG--CTVATVADAVDRGANL-P-AP  
 LLAGTTPERLVVFDYHPPDDDDQRAAFESARRRLADAGSLVIVETLDAVRARGRDL-PAAP  
 VLSGDSVRSVVFDYAEDEDDHRAALESARARL--ADSPVTVDTLDELVARGRDL-PAAP

**Phosphate binding loop ppx(S/T) (S/T/G)G(S/T)TGxPK (A3)**

Firefly\_Luciferase|BAL46512.1  
 FadD13|NP\_217605.1  
 NRPS|WP\_014382786.1  
 Tsukamurella\_paurometabola  
 Mycobacterium\_smegmatis1  
 Mycobacterium\_genavense  
 Mycobacterium\_marinum  
 Nocardia\_asteroides  
 Nocardia\_paucivorans  
 Mycobacterium\_fortuitum  
 Mycobacterium\_smegmatis2  
 Mycobacterium\_vaccae  
 Mycobacterium\_obuense  
 Nocardia\_seriolae  
 Nocardia\_otitidiscaviarum  
 Nocardia\_iowensis  
 Nocardia\_transvalensis

-----ALIMNSSGSTGLPKGVALPHRTA---CVRFSHARDPT-FGNQIIPDTAIL  
 PAVECGGDDNLFIMYTSGTTGHPKGVVHTHESVHSAASSWASTIDVR-YRDRLLLP----  
 A-----DSIAYLIYTSGTTGTPKGAVVTHSNV---TQLESIDAQL-----DVGQVWT  
 FFAEPGTDPLVTLIYTSGTTGTPKGAMYTEQLV---RDRAWLVDSIV---DIDMPAESLL  
 IYTADHDQRLAMLIYTSGTTGAPKGAMYTEAMV---ARLWMTSFIG---DPTPVINV  
 ACADGDDQRLAMIMYTSGTTGTPKGAMYTERTV---TTVWTTMRFLA-----PGLPVINA  
 DCGAQADAPALLIYTSGTTGAPKGVVYTRNRV---ADFWRTSKAEVEATEQRTAPSITL  
 LVPAPQDDPLAMLIYTSGTTGTPKGAMYTDRLV---AAGWQPAR-----PVAVLN  
 LAPAEPPDSRLIYTSGTTGAPKGAMITDRMV---AEHWRAATAER-WGQRGTSPSIVL  
 APKADGSDPLALLIYTSGTTGAPKGAMYLSAV---AKFWRRNSKAW---LGPVSSAINL  
 LYVPDEADPLTLIYTSGTTGTPKGAMYPSKT---ATMWQAGSKAR-WDETGVMPSTITL  
 QF-SPDADTLMLLIYTSGTTGAPKGAMYTERLV---ATTWRRSSRSF-WGDHGLPFSITL  
 PFTSPDDPLLLLIIYTSGTTGAPKGAMYPERLI---TNAWRRSGRSA-WGGEQTTPSITL  
 LHVPADEDEVALIYTSGTTGTPKGAIYPHRLV---TGMWLGPNV-----IPAPVMFM  
 LRIPSDRERLALLIYTSGTTGAPKGAMYTDRLV---AGLWLSANEIR-----VPALTM  
 LFVDTDDPLALLIYTSGTTGTPKGAMYTNRLA---ATMWQGNMSLQ---GNSQRVGINL  
 LHTDGEDELSLLIYTSGTTGTPKGAIYPARLL---TRMWRSGGD-----RMPVVLGF

**possible active site aromatic residue (A4)**

Firefly\_Luciferase|BAL46512.1  
 FadD13|NP\_217605.1  
 NRPS|WP\_014382786.1  
 Tsukamurella\_paurometabola  
 Mycobacterium\_smegmatis1  
 Mycobacterium\_genavense  
 Mycobacterium\_marinum  
 Nocardia\_asteroides  
 Nocardia\_paucivorans  
 Mycobacterium\_fortuitum  
 Mycobacterium\_smegmatis2  
 Mycobacterium\_vaccae  
 Mycobacterium\_obuense  
 Nocardia\_seriolae  
 Nocardia\_otitidiscaviarum  
 Nocardia\_iowensis  
 Nocardia\_transvalensis

SVVPFHHGFMFTTLGYLICGFRVVLMYRFEELFLRSLQDYKIQSALLVPTLFSFF---  
 --LPMFHVAAALTTFVFSAMRGVTLISMPQDFATKVWSLIVEERVCTGGAVPAIINFMRQ-  
 QCHSLAFDFSVWEVFGALLHGGRLVVDVVRSPDDLRLVLRQVSVLSQTPSAFY--  
 HFLPMSHMYGRNWIIAGLASGGTGYFAGASDMSTLFDLLAARPTAIGLVRPVCALIHQR  
 NFEMPLNLHGRIPISTAVQNGGTSYFVPESDMSTLFEDLALVRPTELGLVPRVADMLYQH  
 NFEMPLNLHGRLPLASAFLSGGTSYFVPESDLSTLFEDLALVRPTEAAMVPRVEMLYQH  
 NFMPMSHANGRVLYGTLSNGGTAYFTARSDDLSTLFDLALVRPTELGFPPRIWDMQLER  
 NFLPMSHIAARLTNLGVLAGGTAYFTAADMSTLFDLALVRPTEIFLVRPVCMDLLHR  
 GFMPMSHILGRAICWMALGSGGTVYFAAKSDLSTLDDLALVRPQTTFVPRVWEMLFQR  
 SFMPMSHVMGRGILYASLAAGGTCYFAARSDLSTLLEDLALTRPTELNFVPRVWEMHISE  
 NFMPMSHVMGRGILCSTLASGGTAYFAARSDLSTLFDLALVRPQTLNFVPRVWDMFLQE  
 NFLPMSHVMGRGLLYATLGAGGTAYFAAKSDLSTLFDLALVRPQTLNFSVPRIWDIIFAE  
 NFMPMSHMMGRGVLYGTLGAGGTAYFAARSDLSTLFDLALVRPQTLNFSVPRIWDTIAAE  
 CYMPLSHVAGRMVLSGTFARGGTAYFAASDMSTLFDLALVRPTEVFFVPRVCDMLFQR  
 NYMPLSHIAGRMSLYGTLMRGGTAYFAAASDMSTLDDFGLARPTFLVPRVCELLHQR  
 NYMPLSHIAGRISLFGVLARGGTAYFAAKSDMSTLFDLIGLVRPTEIFFVPRVCDMVFQR  
 SYMPLSHVAGRGLISLALSGLGTVYFAARSDMSTLFDLITLCPMTVFFVPRVCDMVFQR

Firefly\_Luciferase|BAL46512.1  
 FadD13|NP\_217605.1  
 NRPS|WP\_014382786.1  
 Tsukamurella\_paurometabola  
 Mycobacterium\_smegmatis1  
 Mycobacterium\_genavense  
 Mycobacterium\_marinum  
 Nocardia\_asteroides  
 Nocardia\_paucivorans  
 Mycobacterium\_fortuitum  
 Mycobacterium\_smegmatis2  
 Mycobacterium\_vaccae  
 Mycobacterium\_obuense  
 Nocardia\_seriolae  
 Nocardia\_otitidiscaviarum  
 Nocardia\_iowensis  
 Nocardia\_transvalensis

-----AKSTLIDKYDLNLHEIASGG-----APLSKEV-GEAVAK  
 -----VPEFAELDAP-----DFRYFITGG-----APMPEALIKIYAAK  
 -----ALQSADALAPEVGEQLKLQTVVFGGEALE-----PHRL-STWLHH  
 YLA-----VEADTDAET---ARVELRDRVLGGRLQAAMCGSAALSSSEL-QTFMEW  
 HLATVRL-----VTQGADELTAEKQAGAELEQVLGGRVITGFVSTAPLAEM-RAFLDI  
 YRGAVDRG---IAEGADPATAEHDAATEMREQVLGGRLVGGFVGSAPLATEM-KAFLDS  
 FGREVDRLRDGTAEAGDPGALKARVAADLRQVLGGRYALAMMGSAPISEQM-KASVES  
 FRREVDR---ADAGVDEVLAEVVRGELRERVLGGRLTLVLCGSAPIAPEL-RRFVES  
 FQSEMYRR---AADGREQAEVAAEVTAELRRNLGGRLSATTSAPMSAEM-KVWVED  
 YQSRVDQR---LAEGRDREAEVLAEVDRDKVLGGRFVAAMTGSAPISAEL-KTWTQD  
 YQSRDLNR---RAEGSE-DRAEAAVLEEVRTQLLGGRFVSALTGSAPISAEM-KSWVED  
 VAKELERR---TADAAD-----VLADLRQSLGGRYVSAMTGSAPLSAEM-ESFVEQ  
 VAKEVDRR---PDDLAD-----VYADLRQSLGGRHVMAMSGSAPLSP-RTFVED  
 CQSEVQRR---TAAGESVEDADA VKTALREEFLGGRLVVRVMVGSAPVSAEM-KEFMRS  
 YQSELDR---VVGEDAETAATNVKAELRERVLGGRYLTALSGSAPLAEM-KTFMES  
 YQSELDR---SVAGADLDTLDREVKAADLRQNYLGGRLVAVVGSAPLAEM-KTFMES  
 YRSEVDRR---LAAGGDREQVERVKTELREHFLGGRLLLALVGSAPLSP-RAFMES

**active site and mg<sup>2+</sup> binding a(G/W)x(A/T)E (A5)**

Firefly\_Luciferase|BAL46512.1  
 FadD13|NP\_217605.1  
 NRPS|WP\_014382786.1  
 Tsukamurella\_paurometabola  
 Mycobacterium\_smegmatis1  
 Mycobacterium\_genavense

RFHLPGRIGYGLTET---SAILITPEGDDKPG-AVGKVPFFFEAKVVDLDTG---KTLL  
 NIE---VVQGYALTESCGGGTLLSLEDALRKAG-SAGRATMFT---DVAVRGDDGVIREHG  
 HPGLPRMINMYGITE---TVHASFREILRGDNDNVSPIGVPLGNLAFFVLDGWLPRVP  
 LLGID-IQICYGSTEA---GGVIRDGVVVRPPV-TEYKLIDVP---ELGYFTD---SPHP  
 TLGAH-IVDGYGLTET---GAVTRDGVIVRPPV-IDYKLIDVP---ELGYFTD---KPYF  
 MLDH-ITDGYGLTET---GMLTRDNNVNRNRV-IDYKLIDVP---ELGYFLTD---RPYP

Mycobacterium_marinum	LLDLD-VMEGYGSTE	---GTVIINNEVQRPQV-IDYKLVDDVA--ELGYFLTD---RPYP
Nocardia_asteroides	VLRLR-LHDGYGSTE	---GGVIFDTKVMRPPV-LDYKLVDDVP--ELGYFSTD---KPYP
Nocardia_paucivorans	LLDLP-LRDGYGSTE	---GSITLDGRVRRPPV-IDYKLVDDVP--ELGYFSTD---RPYP
Mycobacterium_fortuitum	MLGIH-LLEGYGSTE	---GMALFDGVVQRPV-IDYKLVDDVP--DLGYFGTD---QPHF
Mycobacterium_smegmatis2	LLDMH-LLEGYGSTE	---GAVFIDGQIQRPV-IDYKLVDDVP--DLGYFATD---RPYP
Mycobacterium_vaccae	LLDMH-LIDGYGSTE	---GAVLVDGQIQRPV-IDYKLVDDVP--DLGYFSTD---RPHF
Mycobacterium_obuense	LIDIH-LTDGYGSTE	---GAVFVDGQVQRPV-IDYKLVDDVP--DLGYFTTD---RPHF
Nocardia_seriolae	VMGQP-VIDGYGSTE	---GGILIDNEIRPPV-IDYKLADVP--ELGYFSTD---KPHF
Nocardia_ottitidiscaviarum	LLDDE-LHDGYGSTE	---GSVLLDNRIKRPV-LDYRLVDDVP--ELGYFSTD---KPHF
Nocardia_iowensis	VLDLP-LHDGYGSTE	---GSVLLDNQIQRPV-LDYKLVDDVP--ELGYFSTD---RPHF
Nocardia_transvalensis	VLEIG-MFDGYGATE	---GGVLLNNELQRPV-LDYRLVDDVP--ELGYFGTD---KPYP

Firefly_Luciferase BAL46512.1	Distorted beta sheet GEx <sub>10-14</sub> GY (A6)	ATP binding (S/T)GD (A7)
FadD13 NP_217605.1	VNQRGELCVRGPMIMSGYVNNPEATNALIDKDGW-----LHSGDIAYWDEDEHFFIVD	
NRPS WP_014382786.1	---EGEVVKSIDILLKEYWNRPEATRDADF-NGW-----FRTGDI-GEIDDEGYLYIK	
Tsukamurella_paurometabola	VGWVGELYVAGGGLATGTVGRPLSATRFVACPFAGARMYRTGDLVRSADGQLEYIG	
Mycobacterium_smegmatis1	---RGELLVKSTQLIPGYNSDK---RIRDEGF-----YRTGDMVAELGPDRLVYVD	
Mycobacterium_genavense	---RGELLVRSQTLPFGYKRPVETASVFDGDRGY-----YHTGDMVAETAPDHLVYVD	
Mycobacterium_marinum	---RGELLVKTDTMTPGYKRPVETAQVFDEDDGF-----YKTGDMVAEIEPDHLVYVD	
Nocardia_asteroides	---RGELLVKTRTLFSGYVYRDPEDGQVDFPDGF-----YRTGDMVAQVGPDRLAYLD	
Nocardia_paucivorans	---RGELLTKTTMISGYRRPEVTAQVFDEDDGF-----CRTGDMVAELGPDRAVYVD	
Mycobacterium_fortuitum	---RGELAVRSESLTPGYRRPDVTAQVFDADGY-----YHTGDMVAEIVAPDHLVYVD	
Mycobacterium_smegmatis2	---RGELLIKTENLFPFGYKRPVETASVFDDEDDGF-----YRTGDMVAEIGPDQLVYVD	
Mycobacterium_vaccae	---RGELLVKSEMFPGYKRPVETAEIMFDEDDGY-----YRTGDMVAELGPDHLVYVD	
Mycobacterium_obuense	---RGELLVSESLTPFGYKRPVETAEIMFDEDDGY-----YRTGDMVAETAPDRLTYLD	
Nocardia_seriolae	---RGELLVKSETLFPFGYKRPVETAEIMFDDPGY-----YRTGDMVAETGPDQLVYVD	
Nocardia_ottitidiscaviarum	---RGELLVKSTQQLIPGYFKRPDVTAEIFDADGF-----YRTGDMVAEVRPDHLVYVD	
Nocardia_iowensis	---RGELLTKTESMFPFGYKRPVETAEIMFDDADGF-----YRTGDMVAELGPEQLVYVD	
Nocardia_transvalensis	---RGELLKKAETLIPGYKRPVETAEIFDEDDGF-----YKTGDMVAELEHDLVYVD	
	---RGELLVKSETLVPFGYKRPVETAEIFDADGF-----YRTGDMVAEIGDRDLVYVD	

Firefly_Luciferase BAL46512.1	Hinge domain Rx(D/K) <sub>x</sub> G modified to Rx <sub>4</sub> Kx <sub>3</sub> G in CARs (A8)
FadD13 NP_217605.1	RLKSLIKY-KGYQVAPAELESILLQHPNIFDAGVAGLPDDDAGELPAVVVLEHG----
NRPS WP_014382786.1	DRLKDMIISGGENVYPAEIESVIIIGVPGVSEVAVIGLPDEKWEIAAAIVVAD-----
Tsukamurella_paurometabola	RADEQVKI-RGYRIELGEIRAAADVEGVEQAALVLRDRAGEKRLVGVY-----
Mycobacterium_smegmatis1	RRSNVIKLAQGEFVPIAQLAIAAGPDVHQIFLYGTSESRYS--LIGVVVPAPGPDGET-
Mycobacterium_genavense	RRNNVLKLAQGEFVAVANLEAVFSGAALVRQIFVYGNSESRF--LLAVVVPTPEALEQD-
Mycobacterium_marinum	RRNNVLKLAQGEFVAVANLESIYAGAPLVRQIFVYGNSESRN--LLAVIVPTPEALAEGR-
Nocardia_asteroides	RRNNVLKLSQGEFVAVSLEAIFANSPLVRQIFVYANGARAY--PLAVVVPTQDAQSRHG-
Nocardia_paucivorans	RRNNVLKLSQGEFVTVSRLAFAVAGADLVRQIYVYGSSERAY--LLAVIVPTPEALAGP-
Mycobacterium_fortuitum	RRSFVLKLSQGEFVTVSKLEAVFARSPLVRQIYVYGNSTRSY--LLAVVVPTPEEAQARGD
Mycobacterium_smegmatis2	RRNNVLKLAQGEFVTLAKLEAVFGNSPLVQIYVYGNSAQPY--LLAVVVPTDPSVS---
Mycobacterium_vaccae	RRNNVLKLSQGEFVTVSKLEAVFGDSPLVRQIYVYGNARSY--LLAVVVPTPEEALSRLWD
Mycobacterium_obuense	RRNNVLKLSQGEFVTVSKLEAVFGDSPLVLIHQIYVYGNARSY--LLAVVVPTDAALAR--
Nocardia_seriolae	RRNNVQKLSQGEFVTVSKLEAVFGDSPLVRQIYVYGNARSY--LLAVVVPTEDVLGR--
Nocardia_ottitidiscaviarum	RRNNVLKLSQGEFVAVSKLEAVYATSPLIAQIFVHSGSERSH--LLAVIVPTAARALAL-
Nocardia_iowensis	RRNNVLKLSQGEFVTVAALEAVYATSPLIHQIFVYGNSESRAY--LLAVVVPTDAVLALP-
Nocardia_transvalensis	RRNNVLKLSQGEFVTVAALEAVFASSPLIRQIFVYGNSESRYS--LLAVIVPTDDALGRGD
	RRNNVLKLSQGEFVAVSKLEALYTSPLVRQIYVYGNSESRAY--LLAVIVPTDDAVTAPE

Firefly_Luciferase BAL46512.1	adenylate forming motif Px <sub>4</sub> GKRx(R/K) poorly conserved apart from second active site lysine (A10)
FadD13 NP_217605.1	----KTMTEKEIVDYVAQVTTAKKLRGGVVFVDEVPKGLTGKL---DARKIREILIKAK
NRPS WP_014382786.1	----QNEVSEQQIYVEYCGTRLARYKLPKKVIFAEAI PRNPTGK----ILK---TVLREQY
Tsukamurella_paurometabola	----TGADPSEIRARLGRRLPTVMVPAVVVLDVLP LTVNGKL---DTRALPAPE----
Mycobacterium_smegmatis1	DAQTRTRVLDGLAAIARENDLAAYEVPRDVLIERDPFSQENGLRSGIGKLVRPALIARY
Mycobacterium_genavense	PAALKAAADSLQRTARDAELQSYEPADFLVETEPFSAANGLLSGVGLLRPNLKDRY
Mycobacterium_marinum	SPALKTAIHQSLRQTAAAGLQSYELPVDLFL-ETKPTFDENGLLSGLGKLRLPKLERY
Nocardia_asteroides	RAELKAELHSLRLHRAVAGLAPYEIPRDFIVETTPFTPQNGLLTAIHLKARPHLTQRY
Nocardia_paucivorans	AASLRAELGASLQRAAVAAELEPYEIPRDFLI-ESVPFSDVNGLLSGVGLLRPALKQRY
Mycobacterium_fortuitum	VEALKPLIGRSLRQVAKTAGLQSYEIPRDLIV-DTTPFTLRNGLLTGSRKLARPKLEEHY
Mycobacterium_smegmatis2	----KEAIAESLQEVAREADLQSYEIPRDFIV-ETTPFSLNGLLTGIRKLAWPKLKAHY
Mycobacterium_vaccae	GDELKSRISDSLQDAAARAAGLQSYEIPRDFLV-ETTPFTLENGLLTGIRKLARPKLKAHY
Mycobacterium_obuense	-EDVKTAVAESLQDVARAADLQSYEIPRDFLI-ETTPFTLENGLLTGIRKLARPKLKEY
Nocardia_seriolae	-DDAKALVAESLQNVARAAGLQSYEIPRDFLI-EPTPFTLENGLLTGIRKLARPKLKEY
Nocardia_ottitidiscaviarum	PAERTAAIAESLRQIARDAELESYEIPRDFIV-EDEPFTQENGLLSGIAKLRLPKLRERY
Nocardia_iowensis	AARARAEVSESLQRIAKESGLRPEYIPRDLII-ESEPFTIDNGLLSGIGKLRLPKLKEY
Nocardia_transvalensis	TATLSKALAESIQRIAKDANLQPYEIPRDFLI-ETEPFTIANGLLSGIAKLRLPNLKERY
	PAELKAAIGESLQVARDAELESYEIPRDFIV-EPEAFSMANGLLSGIGKLRLPKLKQRY

Firefly_Luciferase BAL46512.1	KGGKSKL-----
FadD13 NP_217605.1	SATVPK-----
NRPS WP_014382786.1	-----YSVDV-----RYRAPASAEIEILAGIYAQVLGAERVGW--DGSFFD
Tsukamurella_paurometabola	GDRLHDLIYAQADTRQREGRLALDASG---PIIDTVLGAALTLGADIADFDADTRFGD
Mycobacterium_smegmatis1	GQRLEQMYADIAATQANQLRELRRRAATQPVIDTLTQAAATILGT-GSEVASDAHFTD
Mycobacterium_genavense	GEQLERLYSBIAAQVDBIRVLREAADRPAVETLAGACRALLGTSGADS--ESHTFD
Mycobacterium_marinum	GARLELLYTELADSQTRRLHRLRQTGGRLPALETIRRAAGALLGTETTEPRPEAHFKD
Nocardia_asteroides	GARLDALYDEVVREQAELERLRREAPMLPVDEVVARAALAVLGCDAADRPSARFGE
Nocardia_paucivorans	GPRLEQLYTELAEAGTDELRAALRRDGADRPVLETVRRRAVSALLGTTDGAPPEAHFTD
Mycobacterium_fortuitum	GERLEQLYAEALAEQAABELRELRSASADAPVETVSRAGALLGAASDLGPDHFTD
Mycobacterium_smegmatis2	GERLEQLYTDLAEGQANBELRELRRNGADRPVETVSRRAVALLGASVTDLSAHFTD
Mycobacterium_vaccae	GDRLAELYAEALAEQQTDELRELRRSAGERPQETVLRAAALLGAAATDLQPDHFTD

Mycobacterium_obuense	GEQLEALYAEADGQADEMRTLADGANRPMLETVGRAAAALGTAATDVQPDHFTD
Nocardia_seriolae	GARLEQMYDEQAQRQDELATLRREAAELPVLETVCRAARAVLG--GTQPPDAHFTD
Nocardia_otitidiscaviarum	GERLEQLYAEAEQREDELTALRRGAHDRPILDTVTRAAGAVLDLTAGEVSPDAHFD
Nocardia_iowensis	GAQLEQMYTDLATGQADELLALRREAADLPVLETVSRAAKAMLGVASADMRPDHFTD
Nocardia_transvalensis	GDRLEERYDELSREQQDELTAALRTAAADLPVLETVSRAAKALLGCATTDLRPDHFD

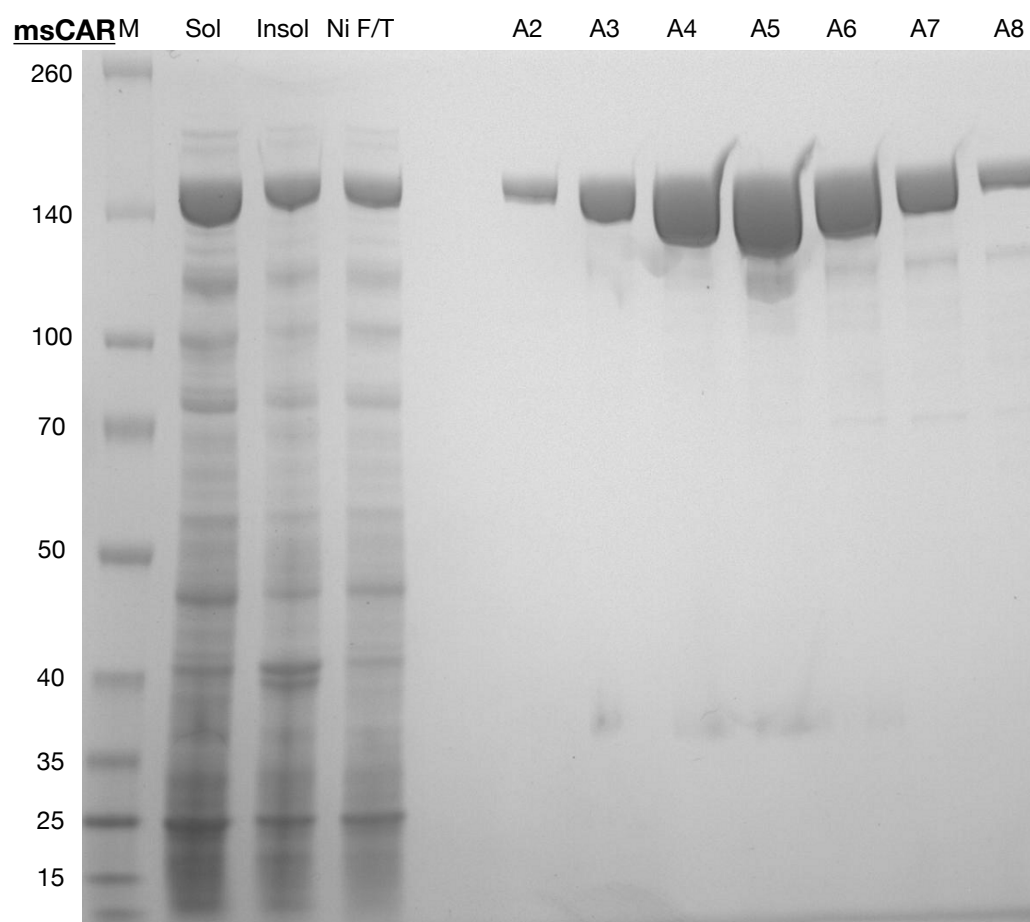
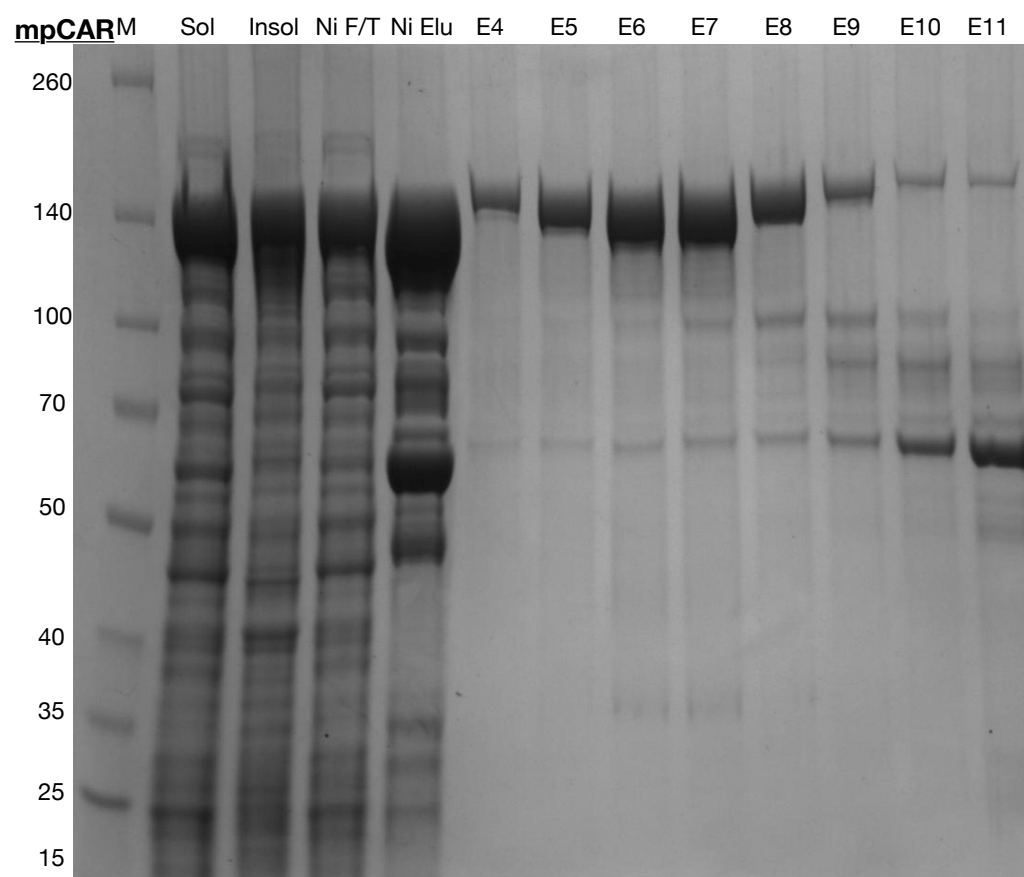
### **Supplementary Figure 4-1 - MUSCLE Alignment of adenylation domains of ANL superfamily members.**

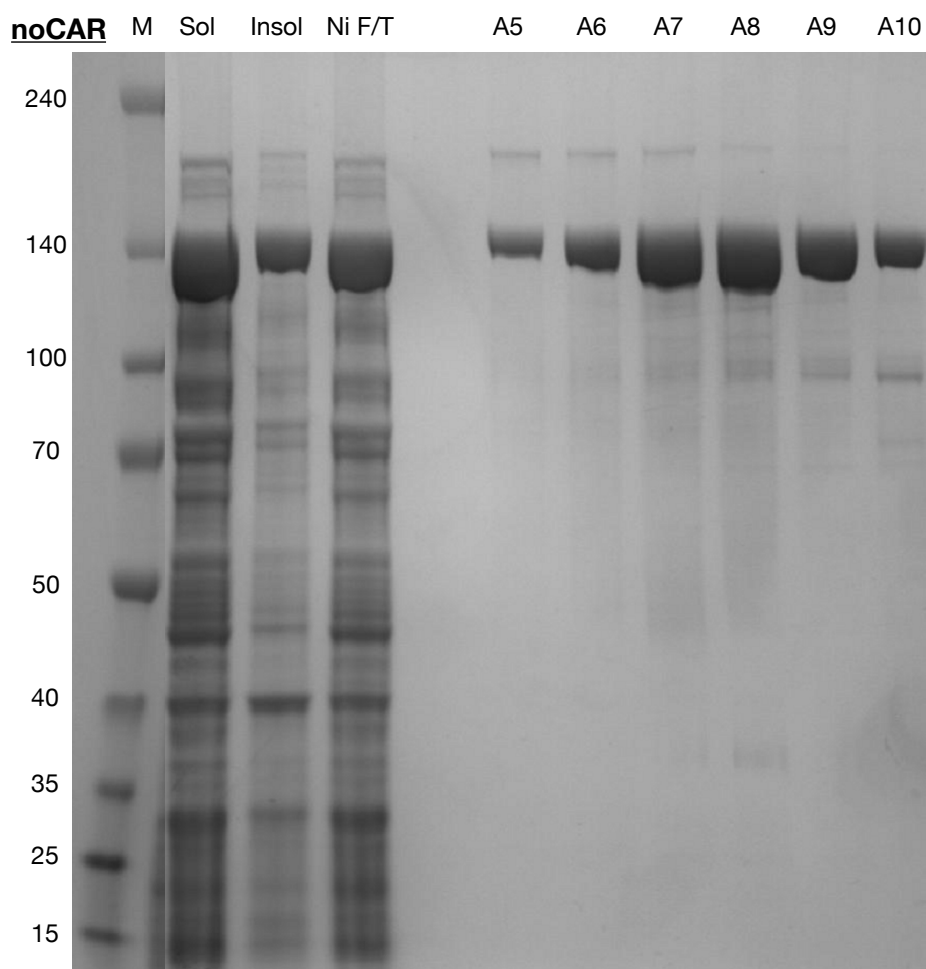
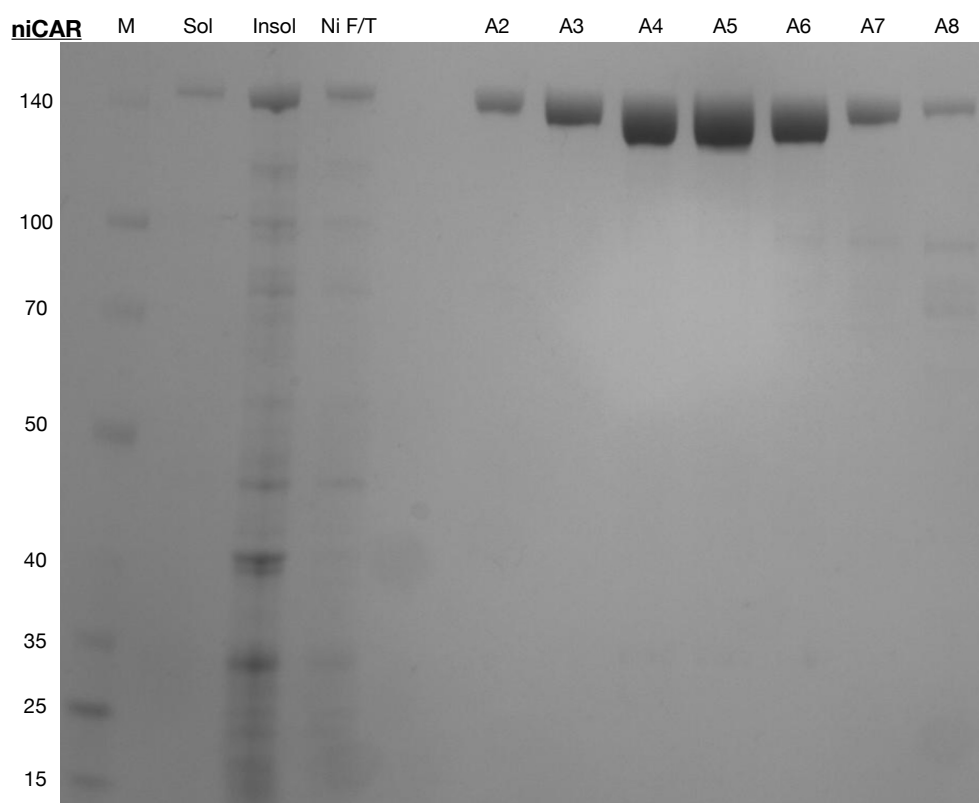
MUSCLE Alignment between Mycobacteria, Tsukamurella and Nocardia CARs, Firefly Luciferase from Photinus pyralis, a long chain fatty acid CoA ligase (FadD13) from Mycobacterium tuberculosis, and an adenyating subunit from a non-ribosomal peptide synthase from Mycobacterium intracellulare suggests the CARs are a member of the ANL enzyme superfamily. Confidently conserved domains that confer the hallmark functions of the ANL family are highlighted blue. Potentially, or poorly conserved domains are highlighted yellow. ANL hallmark motifs were identified according to Gulick 2009 and Marahiel et al., 1997. Accession numbers for the CARs are found in Supplementary Figure 2.

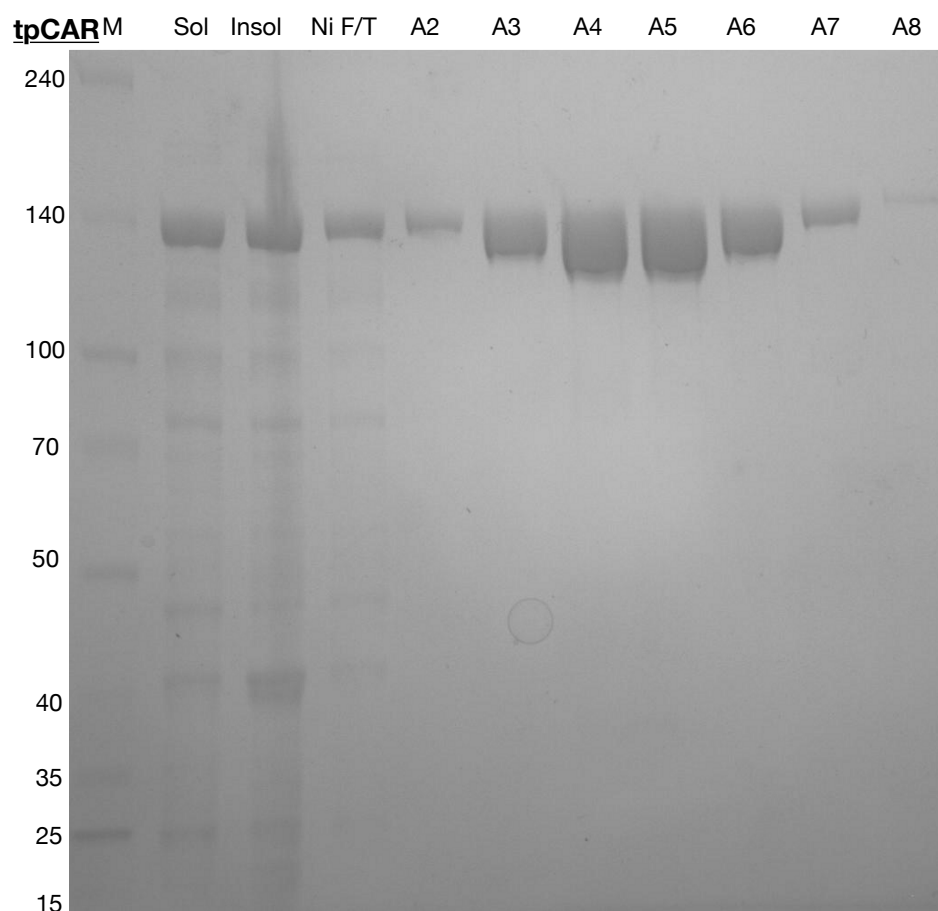
### **CAR Sequences are included with associated data files due to space limitations**

### **Supplementary Figure 4-2 - All CAR sequences used in this study to build phylogenetic tree in figure 2.**

Sequences were retrieved from GenBank by homology search to the *N. iowensis* CAR. It must be noted that the majority of CAR homologues identified are autoannotated as oxidoreductases, and not carboxylic acid reductase.

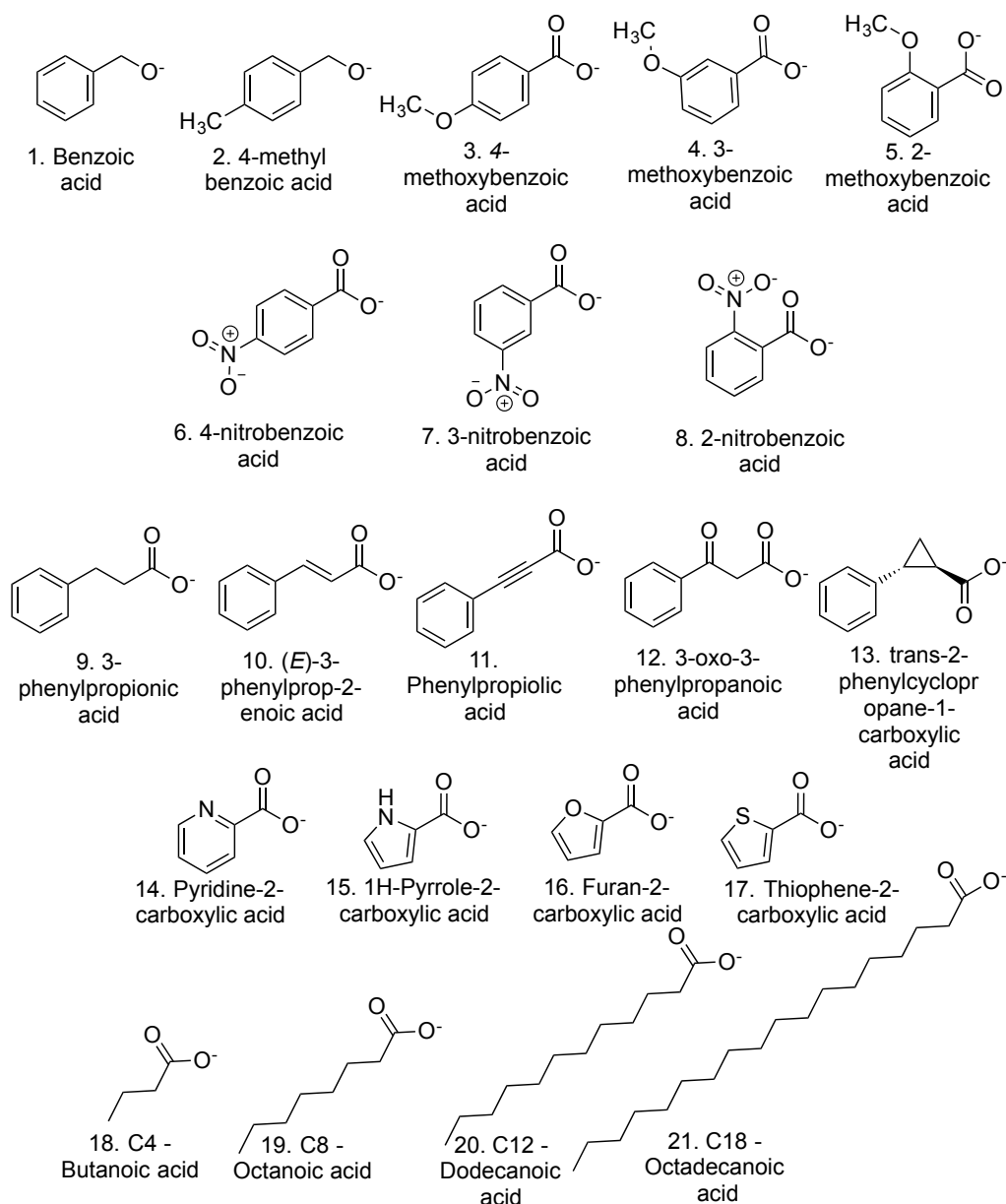






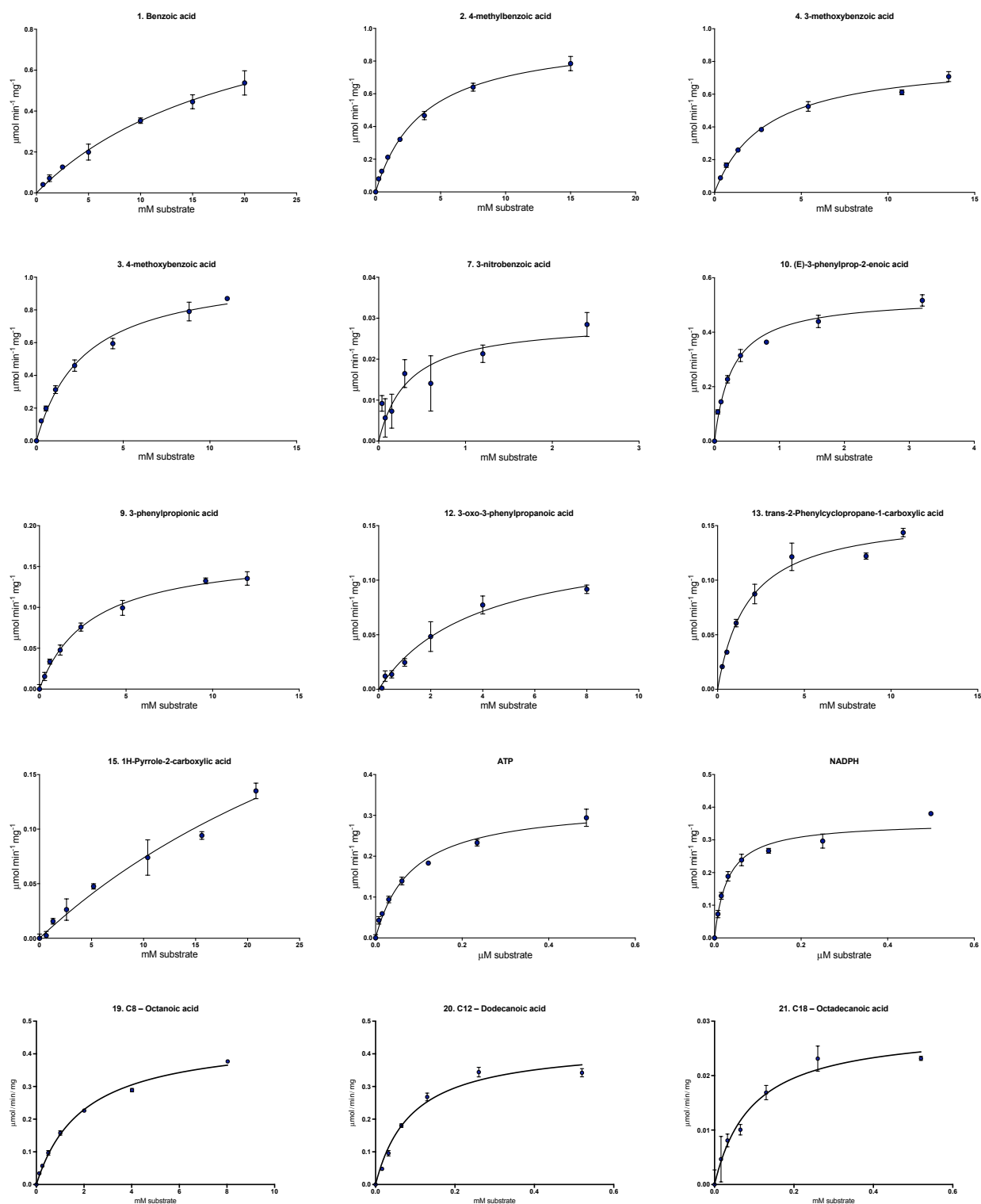
**Supplementary Figure 4-3 – SDS-PAGE analysis of the purification of mpCAR, msCAR, niCAR, noCAR and tpCAR.**

Each gel shows the soluble fraction of the cell lysate (Sol), the insoluble fraction (Insol), the unbound flow-through from the nickel column (Ni F/T) and the fractions which were collected from the subsequent gel filtration column (eg A2-A8). SDS-PAGE analysis of the gel filtration fractions were used to select fractions for pooling for subsequent assays.



#### Supplementary Figure 4-4 – Carboxylic acid substrates

The names and structures of all 20 carboxylic acid substrates tested are shown. The effects of the addition of electron donating or withdrawing groups were tested (2 to 8), extending the carboxylic group away from the benzene ring with various modification (9-13), the inclusion nitrogen, sulfur or oxygen in the benzene ring (14-17) and fatty acids with chain lengths between four and sixteen carbons in length (18-21).

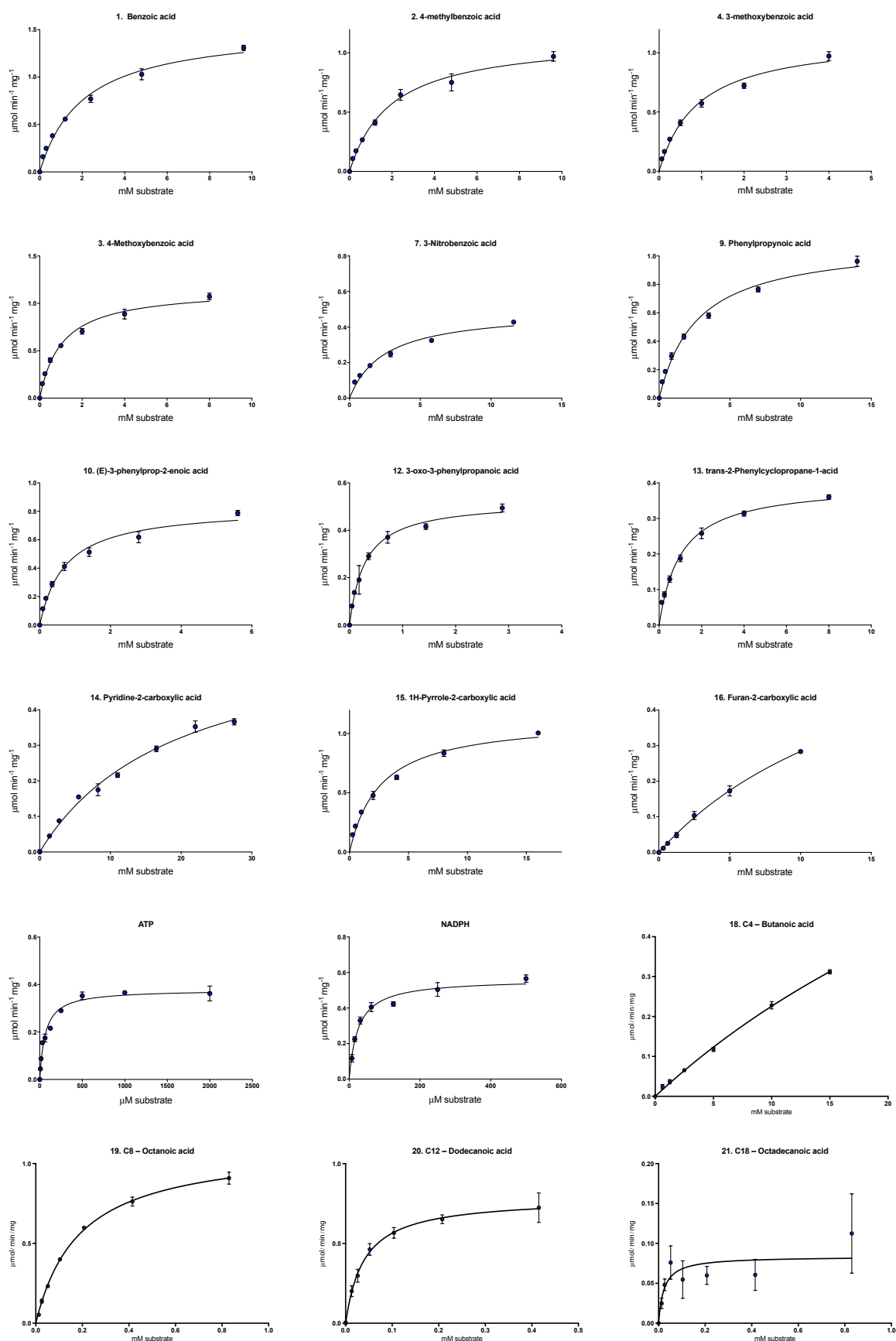


**Supplementary Figure 4-5 – mpCAR kinetic data**

Initial rates of activity at changing substrate concentration for each substrate. Appropriate substrate concentrations were calculated from preliminary experiments determining both which substrates the enzyme showed activity against and a rough estimate of the  $K_M$  value from initial rates of activity over a wide range of substrate concentrations

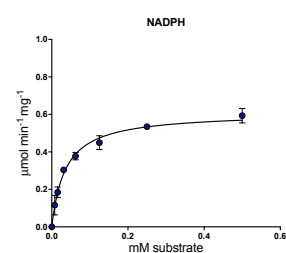
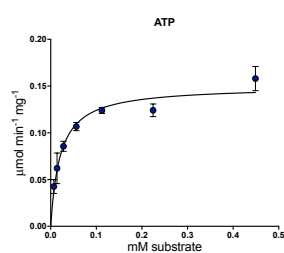
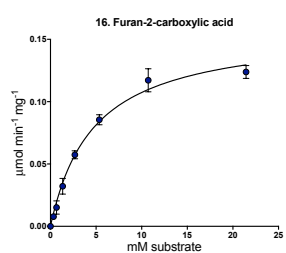
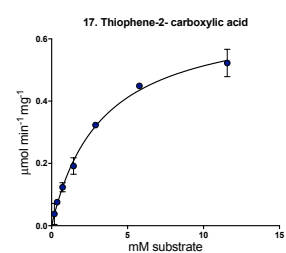
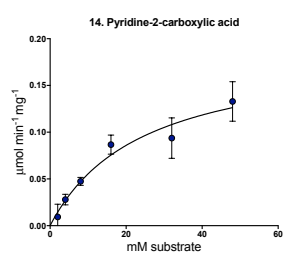
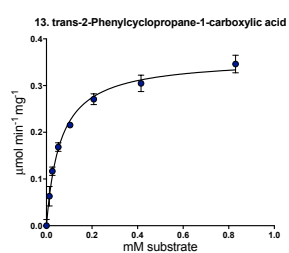
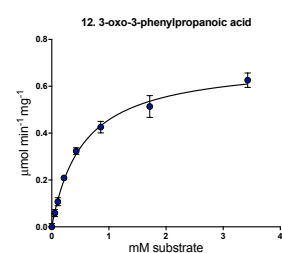
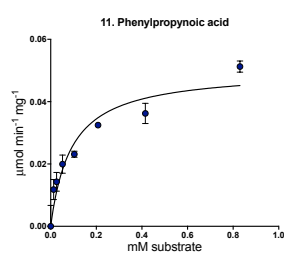
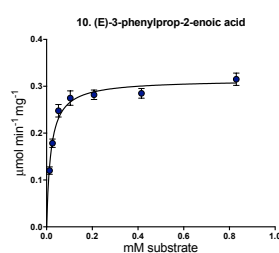
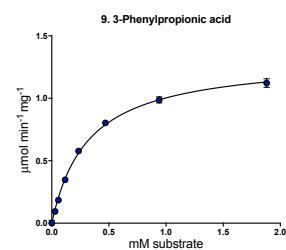
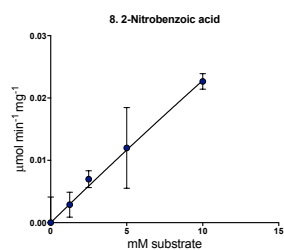
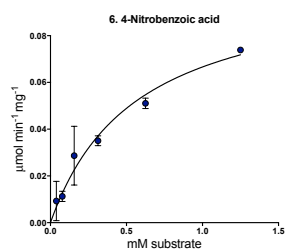
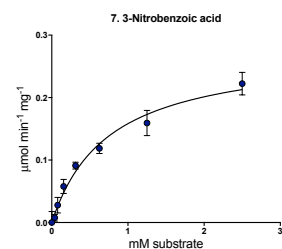
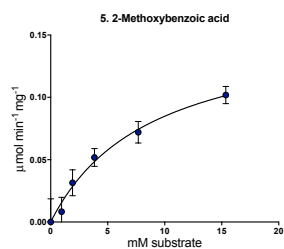
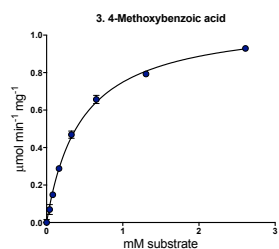
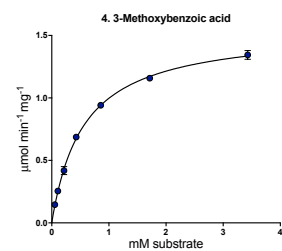
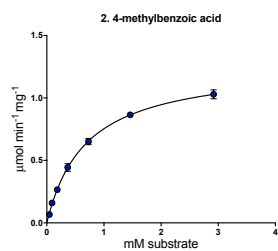
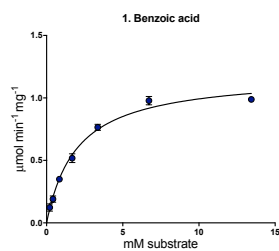


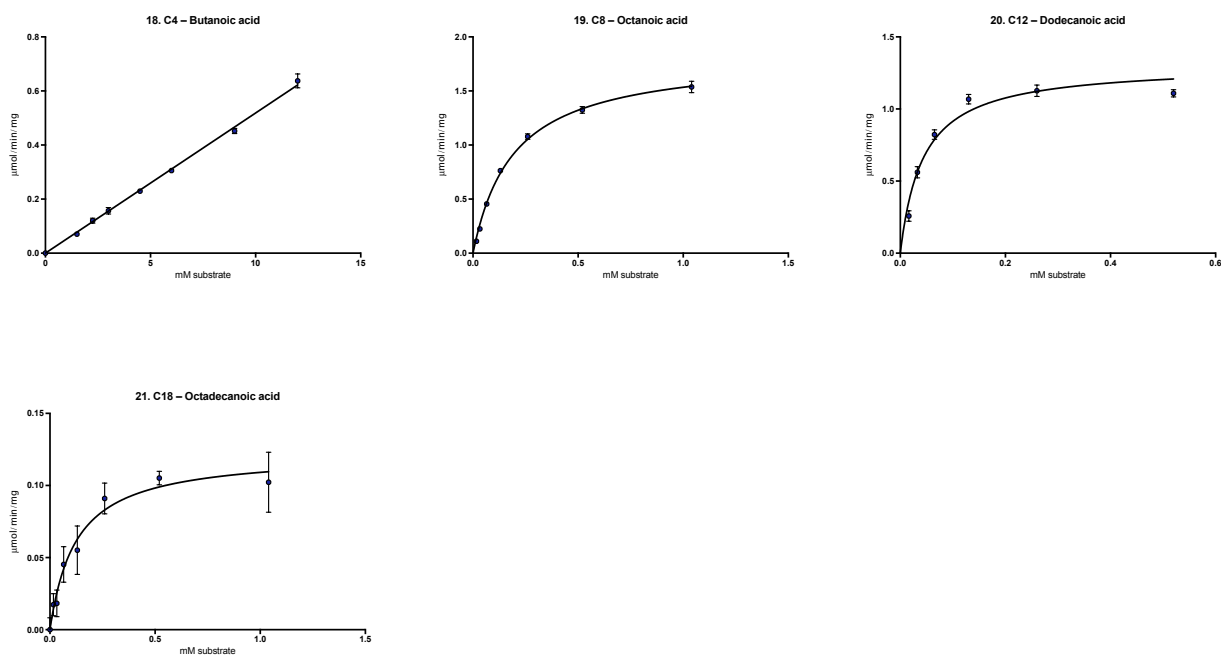




#### **Supplementary Figure 4-6 – noCAR kinetic data**

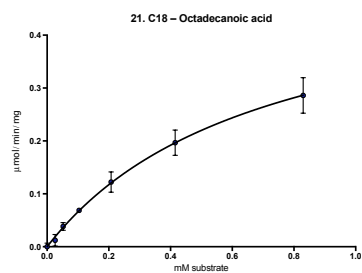
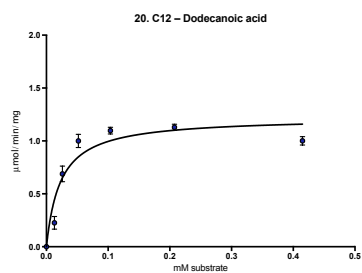
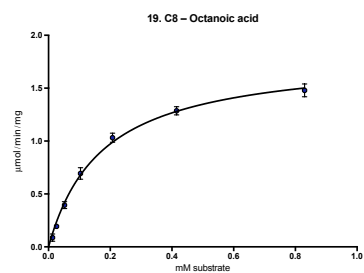
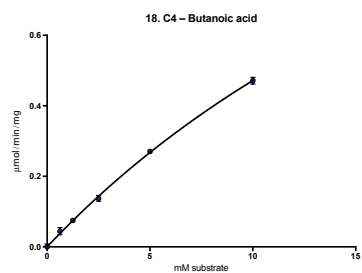
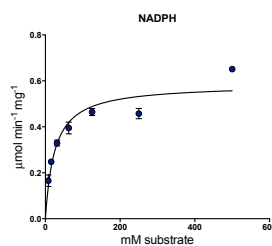
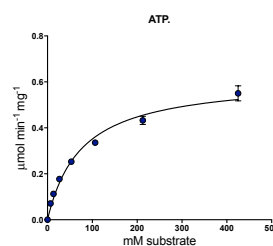
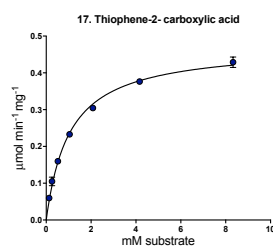
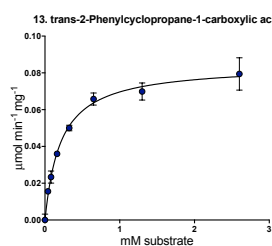
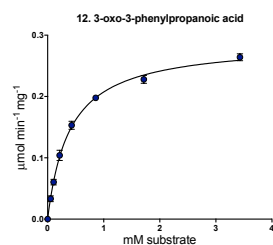
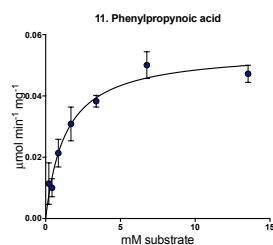
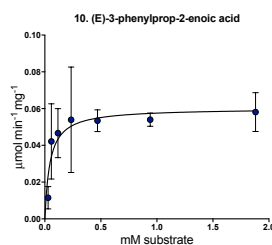
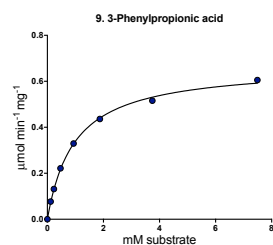
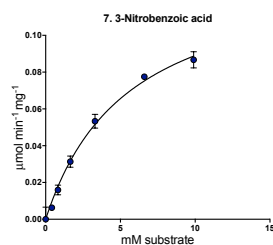
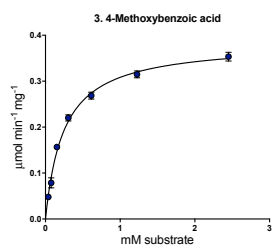
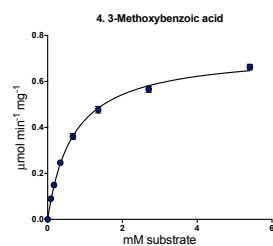
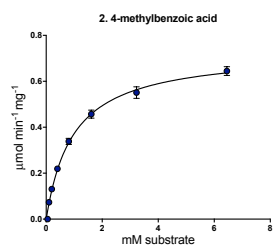
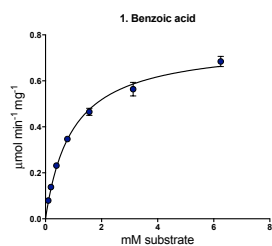
Initial rates of activity at changing substrate concentration for each substrate. Appropriate substrate concentrations were calculated from preliminary experiments determining both which substrates the enzyme showed activity against and a rough estimate of the  $K_M$  value from initial rates of activity over a wide range of substrate concentrations.





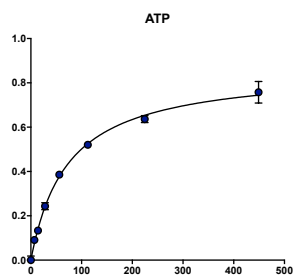
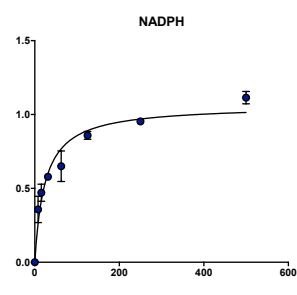
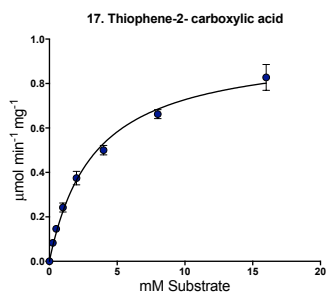
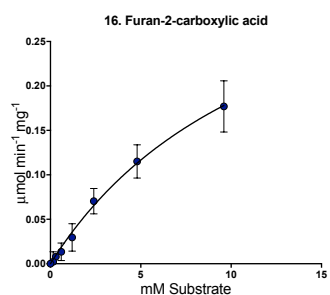
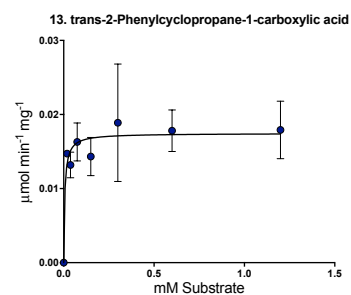
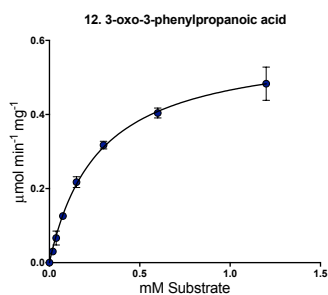
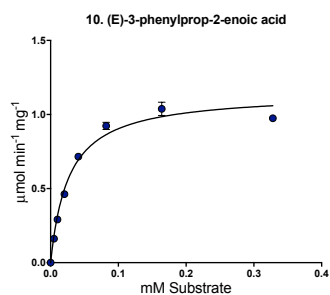
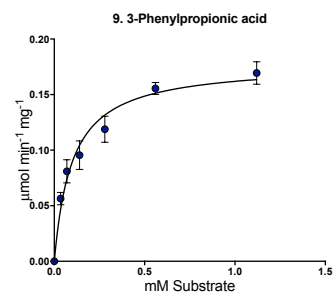
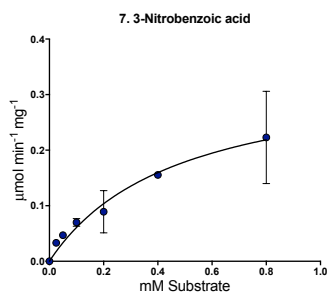
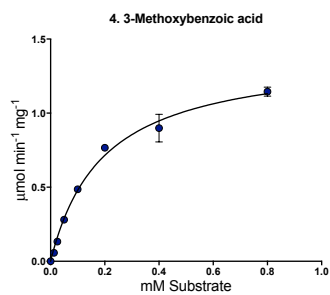
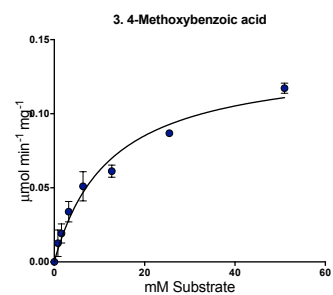
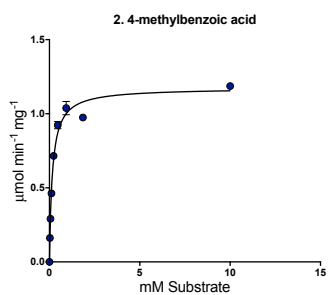
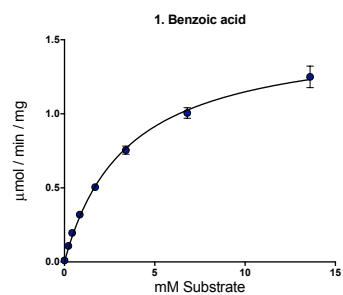
### Supplementary Figure 4-7 – tpCAR kinetic data

Initial rates of activity at changing substrate concentration for each substrate. Appropriate substrate concentrations were calculated from preliminary experiments determining both which substrates the enzyme showed activity against and a rough estimate of the  $K_M$  value from initial rates of activity over a wide range of substrate concentrations.

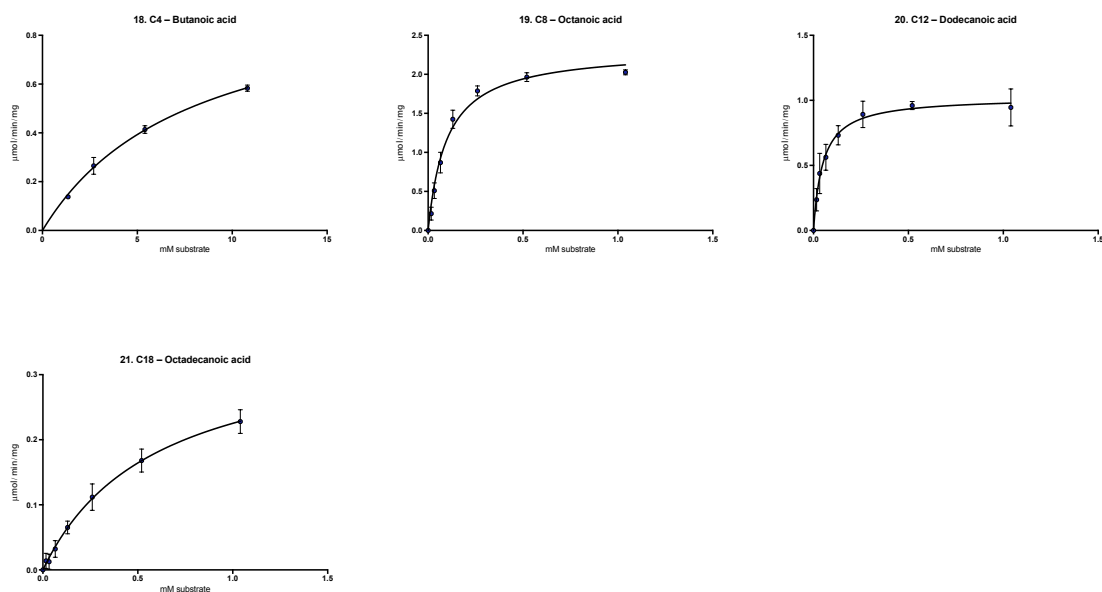


### **Supplementary Figure 4-8– niCAR kinetic data**

Initial rates of activity at changing substrate concentration for each substrate. Appropriate substrate concentrations were calculated from preliminary experiments determining both which substrates the enzyme showed activity against and a rough estimate of the  $K_M$  value from initial rates of activity over a wide range of substrate concentrations.

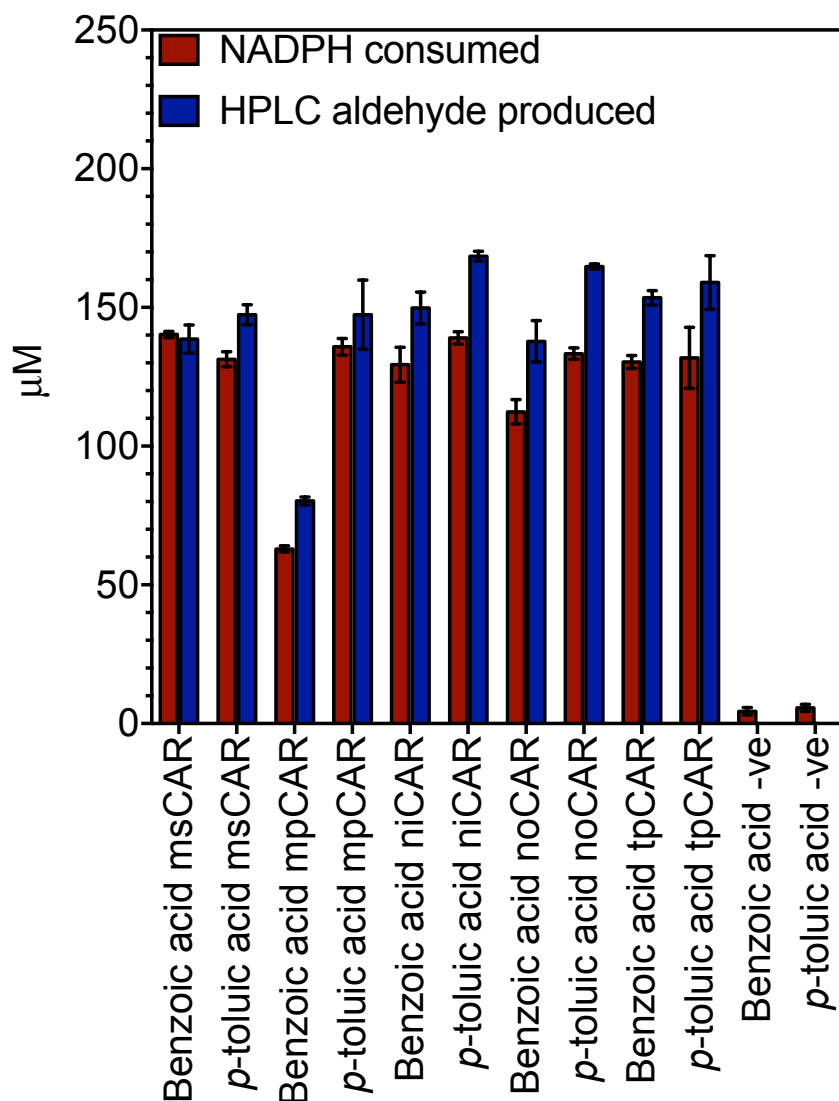






### Supplementary Figure 4-9 – msCAR kinetic data

Initial rates of activity at changing substrate concentration for each substrate. Appropriate substrate concentrations were calculated from preliminary experiments determining both which substrates the enzyme showed activity against and a rough estimate of the  $K_M$  value from initial rates of activity over a wide range of substrate concentrations.

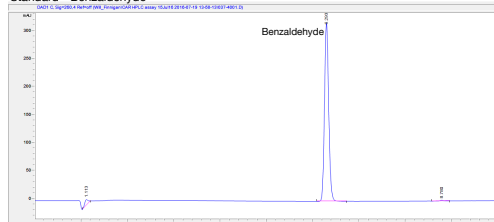


**Supplementary Figure 4-10 – Comparison of NADPH consumption with aldehyde production**

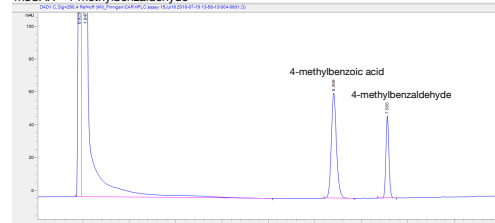
Reactions containing 2 mM benzoic acid or *p*-toluic acid were set up and the consumption of NADPH monitored over 10 minutes at OD<sub>340nm</sub>. Reactions were stopped by the addition of 50 % acetonitrile and aldehyde concentration determined by HPLC. The comparison shows slightly more aldehyde was produced in the reaction than the observed NADPH consumption. This can be attributed to the reaction occurring before the monitoring of NADPH could begin. This comparison shows that measuring NADPH consumption is a good measure of the CAR reaction.



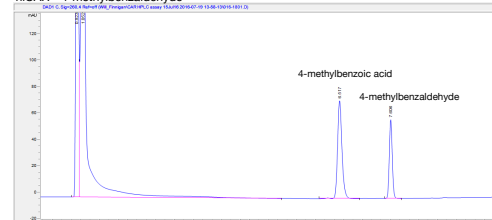
Standard - Benzaldehyde



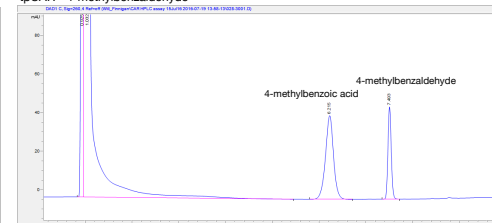
msCAR - 4-methylbenzaldehyde



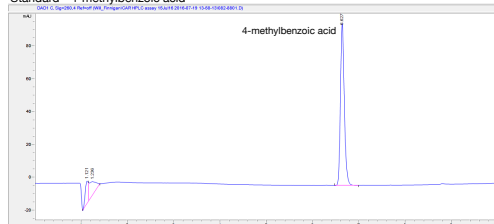
niCAR - 4-methylbenzaldehyde



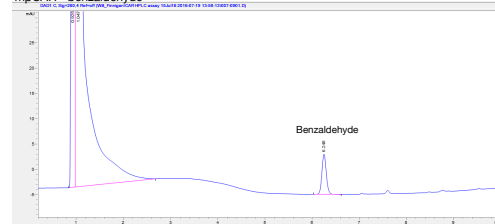
tpCAR - 4-methylbenzaldehyde



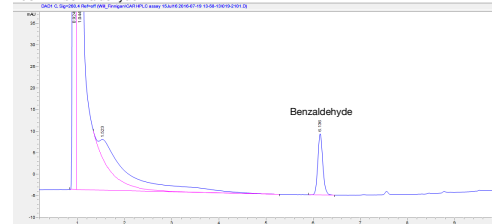
Standard - 4-methylbenzoic acid



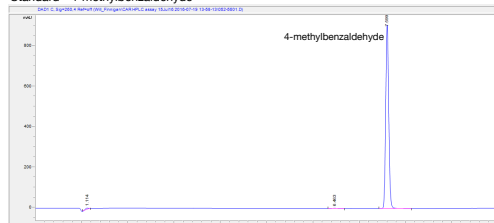
mpCAR - benzaldehyde



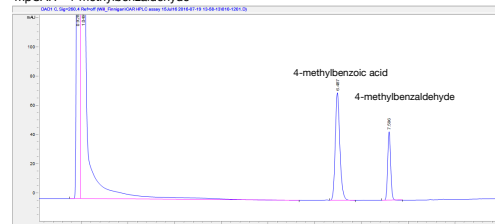
noCAR - benzaldehyde



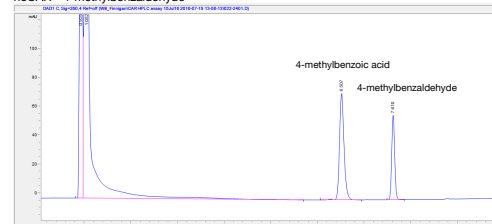
Standard - 4-methylbenzaldehyde



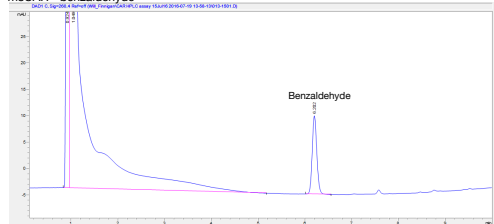
mpCAR - 4-methylbenzaldehyde



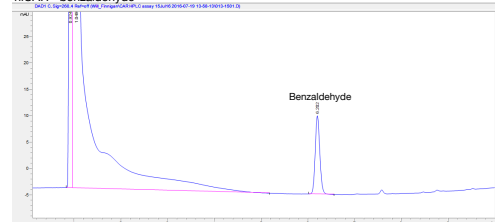
noCAR - 4-methylbenzaldehyde



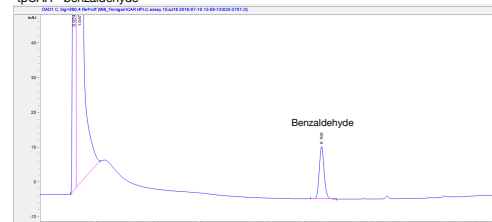
msCAR - benzaldehyde



niCAR - benzaldehyde

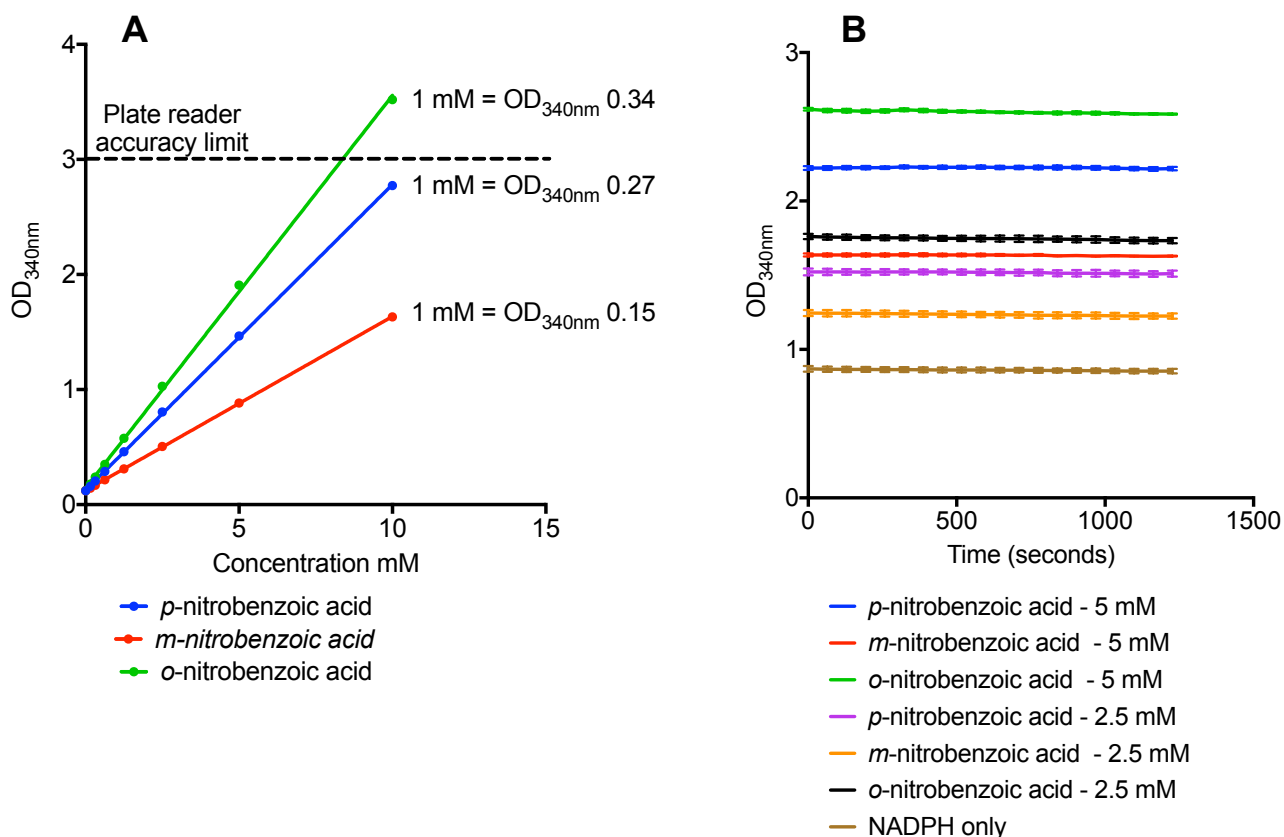


tpCAR - benzaldehyde



**Supplementary Figure 4-11 – Confirmation of benzaldehyde and *p*-tolualdehyde production by all the CARs tested.**

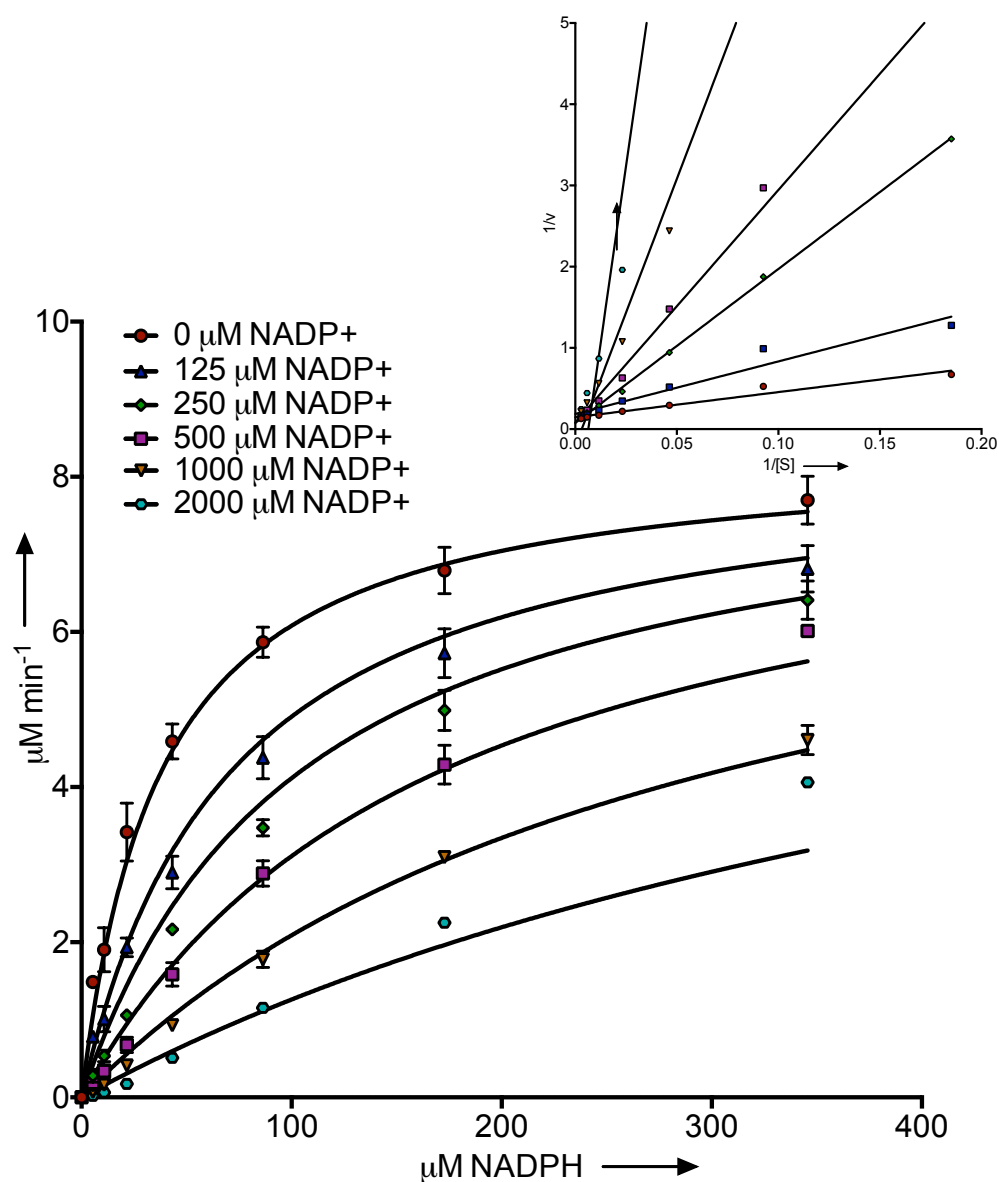
Reactions containing 2 mM benzoic acid or *p*-toluic acid were set up and allowed to run for 10 minutes before being stopped by the addition of 50 % acetonitrile for analysis by HPLC. For all the CARs, the derivative aldehyde product was detected by HPLC.



**Supplementary Figure 4-12 – Controls looking at the feasibility of using NADPH consumption to measure activity of CAR enzymes with nitro aromatic acid substrates.**

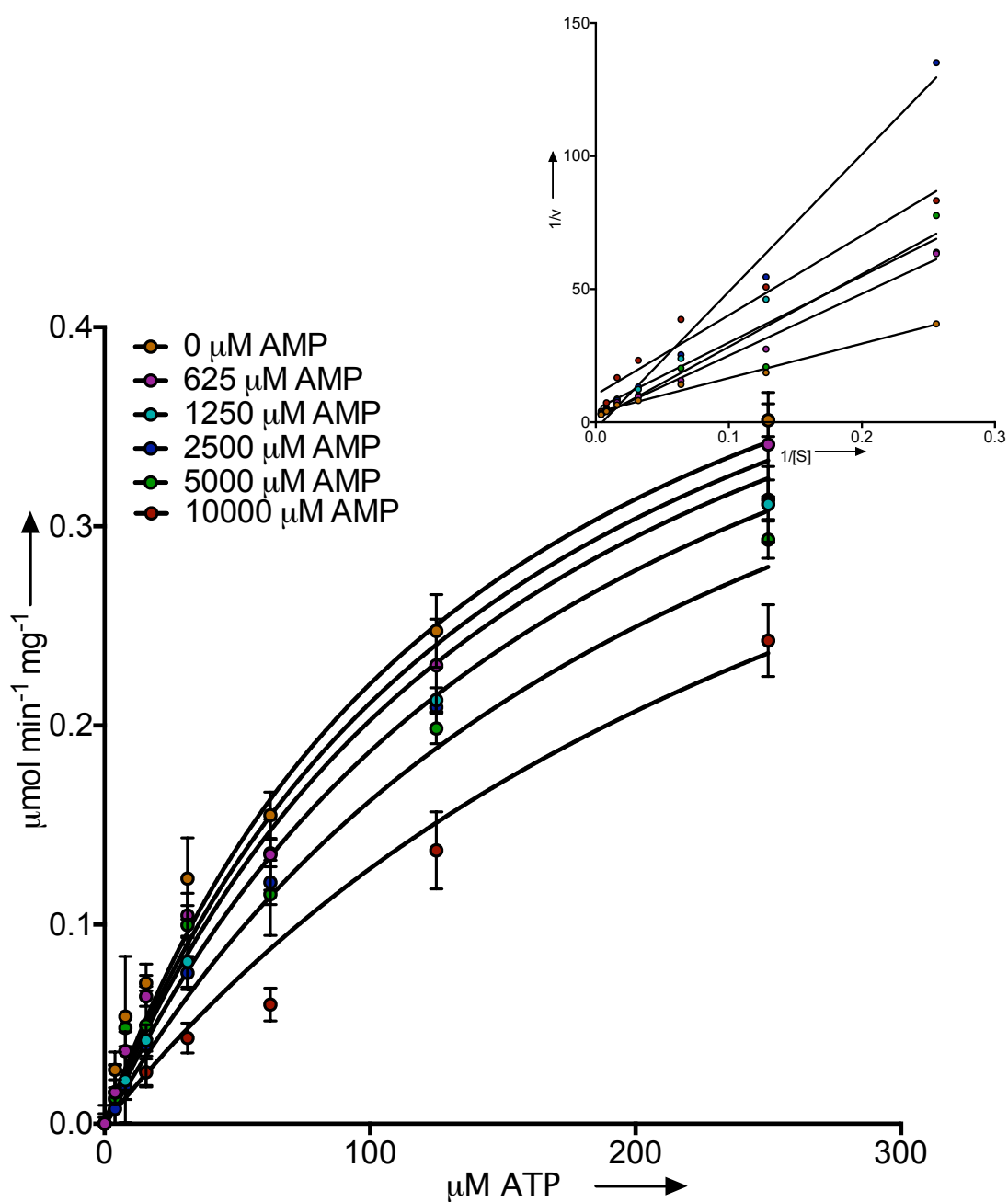
Substrate concentration was titrated from 10 mM and OD<sub>340nm</sub> plotted as a function of concentration. The plate reader used is accurate up to OD<sub>340nm</sub> 3.0 as shown.

Standard CAR reaction mix without CAR enzyme was assayed with the addition of 5 mM and 2.5 mM of the three nitro aromatic acids used. NADPH can be measured accurately as additional absorbance despite the inclusion of the nitro aromatic compounds. Error bars show the standard deviation of three readings.



**Supplementary Figure 4-13 – Competitive inhibition of NADPH by NADP<sup>+</sup>**

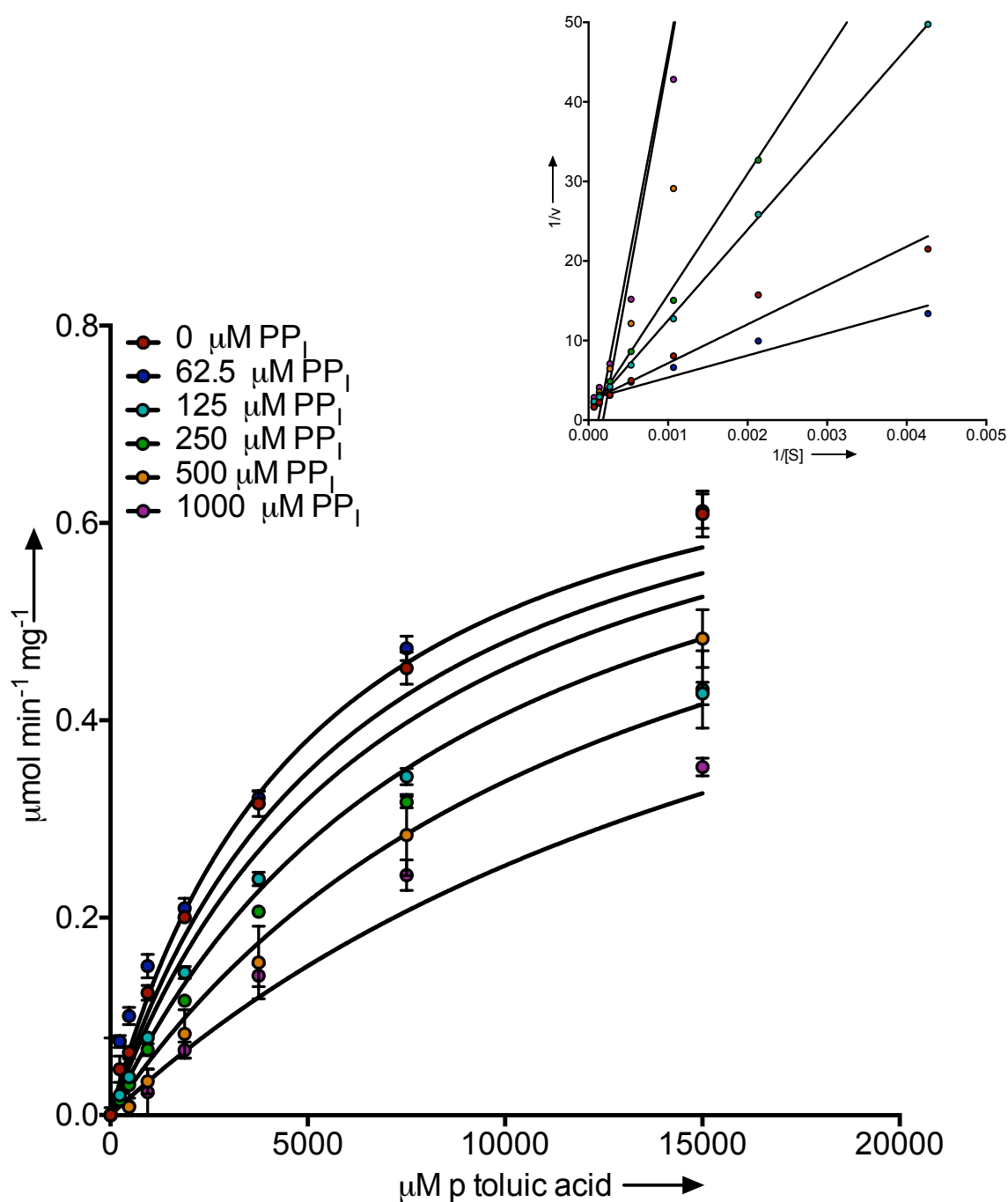
Initial rates at changing concentrations of NADPH at various NADP<sup>+</sup> concentrations were fit to Michaelis-Menten inhibition equations in Graphpad using non-linear least squares regression. The best fit was shown to be competitive inhibition and the Lineweaver–Burk plot in the top right of the figure clearly shows this. The  $K_i$  was calculated as  $143 \pm 8 \mu\text{M}$ .



**Supplementary Figure 4-14 – Competitive inhibition of ATP by AMP**

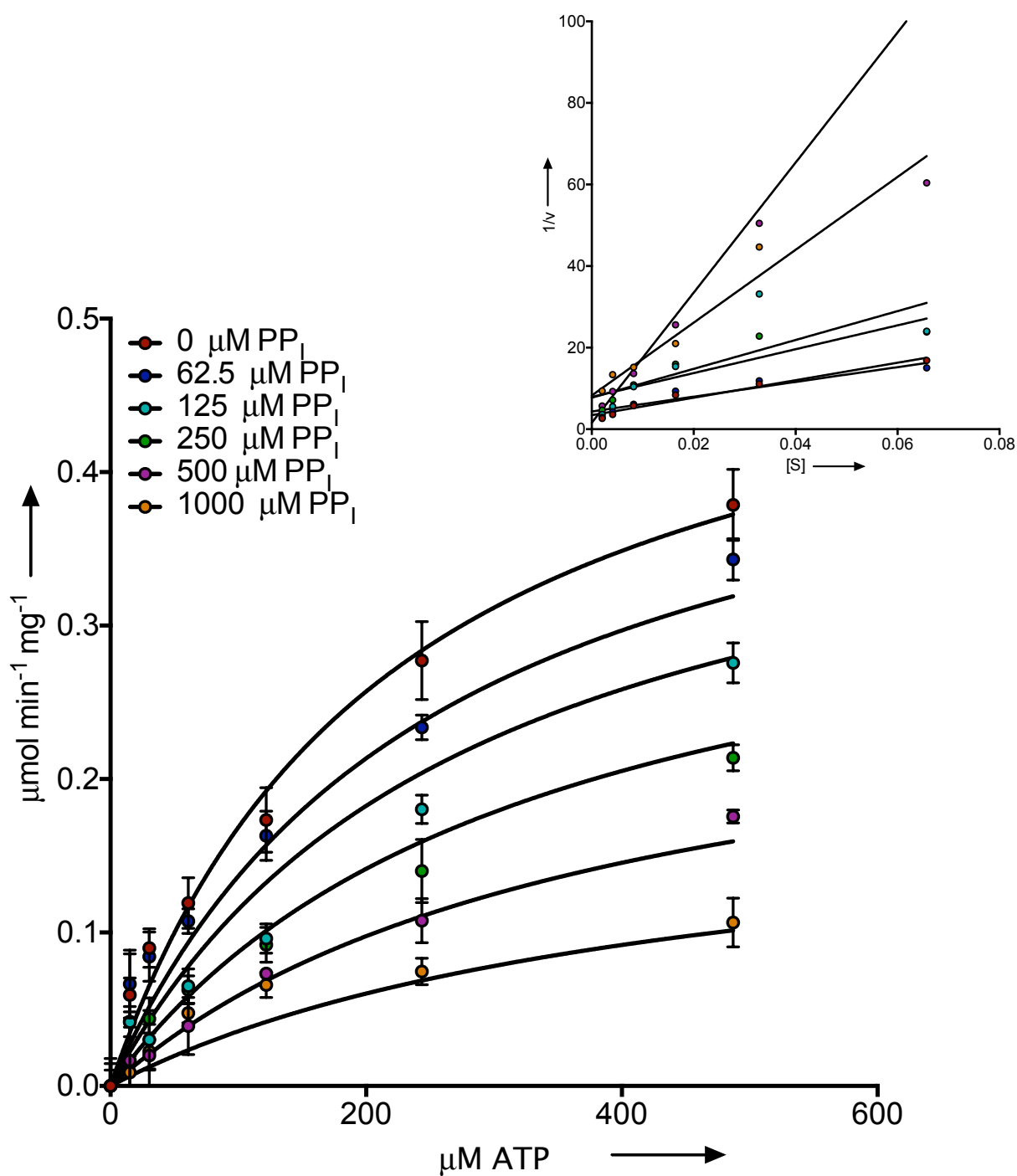
Initial rates at changing concentrations of ATP at various AMP concentrations were fit to Michaelis-Menten inhibition equations in Graphpad using non-linear least squares regression. The best fit was shown to be competitive inhibition. The  $K_i$  was calculated as  $8200 \pm 900 \mu\text{M}$ .





**Supplementary Figure 4-15 – Competitive inhibition of p-toluic acid by PPI**

Initial rates at changing concentrations of p-toluic acid at various PPI concentrations were fit to Michaelis-Menten inhibition equations in Graphpad using non-linear least squares regression. The best fit was shown to be competitive inhibition. The  $K_i$  was calculated as  $340 \pm 40 \mu\text{M}$ .



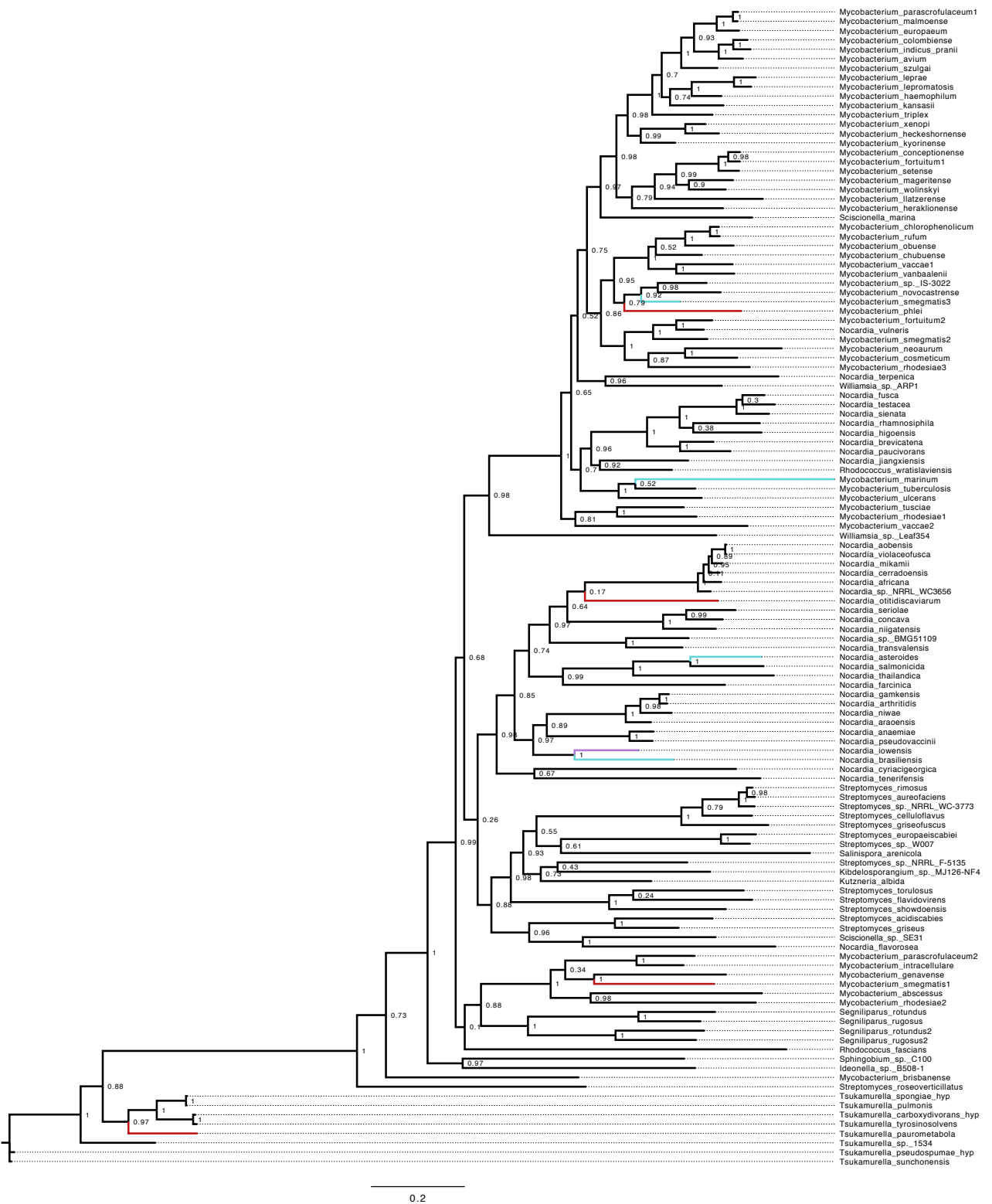
**Supplementary Figure 4-16 – Mixed model inhibition of ATP acid by PPI**

Initial rates at changing concentrations of p-toluic acid at various PPI concentrations were fit to Michaelis-Menten inhibition equations in Graphpad using non-linear least squares regression. The best fit was shown to be mixed inhibition, with a  $K_i$  of  $220 \pm 50$  μM, and an  $\alpha$  of  $2.5 \pm 1.4$ .

**FASTA file is included with associated data files due to space limitations**

**Supplementary Figure 4-17 – All 99 CAR sequences identified**

A FASTA file containing all 124 CAR protein sequences identified. A subset were used for the tree in the main text. An extended tree is shown in supplementary figure 15.



**Supplementary Figure 4-18 - Unrooted maximum likelihood phylogeny of the 124 CAR homologues.**

CARs were aligned with MUSCLE, in Geneious, masked by eye, and had their phylogeny constructed in the PhyML plug-in for Geneious under the WAG+I+G model of amino acid substitution with 8 gamma rate categories. Node confidence scores are SH-like statistics computed by PhyML, with 1 being unequivocal. Coloured branches represent CARs that have been studied: Blue – in previous research, Red – in this paper, Purple – in both this paper and previous research.

	<b>Molecular Weight</b>	<b>Extinction Coefficient</b>
mpCAR	129534.6	116800
msCAR	131391.8	102235
tpCAR	122207.3	108430
niCAR	129519.6	102485
noCAR	129087.2	114850

**Supplementary Figure 4-19 – Table of molecular weights and extinction coefficients calculated using ExPASy ProtParam.**



## **Chapter 5 - Construction, modelling and optimisation of a seven enzyme, one-pot biotransformation**

## **5.1 Preface for inclusion in this thesis**

This chapter is the first draft of a paper intended for submission to eLife. The draft in this thesis is likely an extended version of what will be submitted, containing extra background information in the introduction and more discussion.

All experiments and modelling were performed by William Finnigan. All content written by William Finnigan and edited by Nicholas Harmer.

The manuscript has been formatted to match the formatting of this thesis.

References have been combined with those in the other chapters of this thesis for a final reference chapter. Figure and table numbers have also been edited for inclusion in the thesis.

eLife asks that supplementary figures are linked to figures in the text, and the labelling of the supplementary figures therefore follows this format. For example Figure 5-2 has a number of supplementary figures associated with it labelled Figure 5-2 – Supplementary Figure 1 through to Figure 5-2 – Supplementary Figure 8.



## 5.2 Introduction

### 5.2.1 Multi-step biocatalysis

Biocatalysis, the use of isolated enzymes or whole cells, is increasingly the route of choice in the chemical and particularly the pharmaceutical industries (Pollard and Woodley, 2007). Driven by a requirement for the sustainable use of limited resources, enzymes offer unrivalled selectivity at mild operating conditions making them an excellent choice in meeting these goals (Fessner, 2015).

Recently, there has been an increasing interest in developing one-pot reactions, where multiple catalysts work together to complete chemical pathways. These remove the need to isolate substrates and products at each step, giving great cost and productivity benefits (Turner and Truppo, 2013). The use of multistep pathways can also drive reversible processes to completion and allow inhibitory or unproductive by-products to be eliminated (Oroz-Guinea and García-Junceda, 2013).

Enzymes are particularly well suited to multi-step reactions in that this mimics the use of enzymes in nature, where they carry out complex metabolic transformations in similar and mild conditions. In contrast, the use of several chemical steps together will often have differing operational requirements, making this a more challenging approach (Wells and Meyer, 2014). There is particular interest in implementing enzymatic pathways for the conversion of simple and inexpensive starting materials into high value, often enantiomerically pure, products (Turner and Truppo, 2013).

The use of whole cells for these types of reactions has a number of advantages. These include a simpler upstream process, higher enzyme stability inside the cell, cofactor regeneration utilising host metabolic machinery, and avoiding enzyme isolation steps (De Carvalho, 2011; Schrewe *et al.*, 2013). However, whilst the use of whole cells can reduce the complexity of the upstream process, it suffers from more a difficult downstream process. Moreover, maximising the conversion can be challenging. Diffusion across cell membranes can be rate limiting (Ni and Chen, 2004), and unwanted side reactions can occur (Stuermer *et al.*, 2007). High substrate concentrations can be difficult to achieve due to toxicity to the cells (De Bont, 1998). Finally, the reaction itself is more difficult to control as changes in

enzyme concentration must be engineered into the microorganism, and this is frequently challenging (Oroz-Guinea and García-Junceda, 2013).

The use of isolated enzymes rather than whole cells can significantly reduce the cost and complexity of downstream processing (du Preez *et al.*, 2015). As fewer catalysts and chemicals are present, fewer side reactions occur, and a higher purity product can be achieved (Bujara *et al.*, 2010). Furthermore, reactions are not limited by diffusion across a cell membrane (Hold and Panke, 2009). Perhaps the most attractive feature of using isolated enzymes is the ability to finely control the reaction by changing the concentration of enzymes, substrate, co-solvent, pH or temperature. These allow the design of the most efficient system to meet the process need (Xue and Woodley, 2012). To achieve this, the enzymes involved must be thoroughly characterised, the isolated enzymes must be sufficiently stable for the process to be economical, and cofactor regeneration systems may need to be employed (Oroz-Guinea and García-Junceda, 2013).

### **5.2.2 The enzyme toolbox for *in vitro* biocatalysis**

The concept of an enzyme toolbox is often invoked, together with a need for more well-defined enzyme building blocks (Wells and Meyer, 2014). This correlates with the frequently lamented need for well characterised parts in synthetic biology (where the ultimate goal is the *in silico* design of novel microorganisms) (Kuhn *et al.*, 2010; Ellis, Adie and Baldwin, 2011; Medema *et al.*, 2012). In synthetic biology, modelling of these parts is significantly complicated by the many confounding cellular factors of the cell (Purcell *et al.*, 2013). In contrast, modelling of enzymes *in vitro* can be fairly robust and offers solutions for reaction engineering (Vasić-Rački, Findrik and Vrsalović Presečki, 2011). In order for a robust enzyme toolbox to be realised, well characterised and validated parts are needed (Wachtmeister *et al.*, 2014), ideally with validated mathematical models so that new pathways can be engineered and tested *in silico* (Turner and Truppo, 2013).

For enzymes to be used as part of an *in vitro* system, operational stability is a key factor (Polizzi *et al.*, 2007). Thermostable enzymes are an excellent starting point in achieving high operational stability (Littlechild, 2011). Additional modifications such as enzyme immobilisation or enzyme engineering can be used to further increase enzyme lifespan (Keasling, 2012). The use of thermostable enzymes also allows

purification by heat precipitation: recently a cocktail of nine thermostable enzymes were isolated in this way from a single *E. coli* strain, allowing a convenient one-pot biotransformation (Ninh *et al.*, 2015).

### **5.2.3 Carboxylic acid reductases join up separate enzyme reactions.**

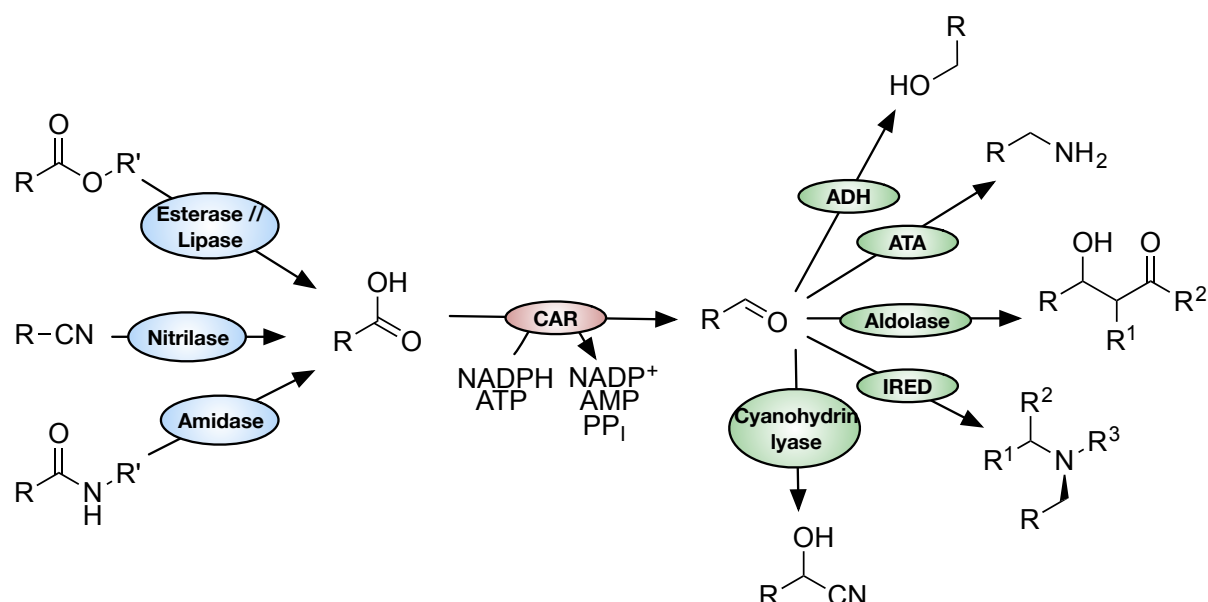
Many critical chemicals for modern society are derived from reduced petrochemical based precursors. The oxidation of these molecules is more chemically straightforward than reduction of the alternative, oxidised sustainable feedstocks. A very relevant challenge is therefore the biocatalytic reduction of renewable feedstocks to provide useful products (Wells and Meyer, 2014). This will considerably assist in the move to a sustainable, bio-based economy whilst avoiding the use to potentially toxic reductants (Cherubini, 2010).

An example of a recent addition to the enzyme toolbox for such reactions are the carboxylic acid reductases (CARs) (Akhtar, Turner and Jones, 2013). CARs catalyse the reduction of carboxylic acids to aldehydes in mild conditions, and can connect many types of enzyme reaction allowing the construction of novel multi enzyme pathways (Figure 5-1) (Turner and O'Reilly, 2013). Previously, we have shown CARs to be a fairly promiscuous class of enzyme, catalysing the reduction of both aliphatic and aromatic carboxylic acids. Electron rich acids are favoured as the first step in the CAR reaction mechanism, attack by the carboxylate on the  $\alpha$ -phosphate of ATP, is limiting (Finnigan *et al.*, 2016).

Hydrolase enzymes are the most widely used class of enzyme industrially (Meyer *et al.*, 2013), generally producing a carboxylic acid as one of their products (Panda and Gowrishankar, 2005). These enzymes are well poised for inclusion into novel enzymatic cascade reactions with CARs. Additionally many enzymes considered as established members of the enzyme toolbox accept aldehydes as their substrates (Turner and O'Reilly, 2013). The use of CARs for multistep reactions opens up a large number of useful enzyme combinations for synthetic chemistry (Figure 5-1).

CARs have been shown to be useful in the construction of pathways for *in vivo* use. Examples include the production of the flavor vanillin by yeast (Hansen *et al.*, 2009) and a synthetic pathway for the production of propane in *Escherichia coli* (Kallio *et*

*al.*, 2014). The use of a CAR in *E. coli* as a whole-cell biocatalyst for the synthesis of the antioxidant 3-hydroxytyrosol has been explored (Napora-Wijata *et al.*, 2014). As well as having beneficial effects on human health it is a valuable building block in the synthesis of pharmaceuticals (Allouche, Fki and Sayadi, 2004). Using a “design of experiments” approach, citrate was shown to be a suitable co-substrate for the regeneration of NADPH and ATP. However, a preparative-scale bio-reduction gave much lower yields than expected, possibly due to mechanical stress or oxygen limitation at the larger scale (Napora-Wijata *et al.*, 2014). The use of CARs as an isolated enzyme, however, has not been explored, possibly due to the challenging and costly requirements for both ATP and NADPH regeneration (Zhang, 2015).



**Figure 5-1 – Potential for utilising CARs in multi-step enzyme reactions**

CARs join many industrially relevant enzyme reactions, making them useful for the construction of novel multi-step enzyme reactions. Cofactors and additional substrates are not shown for clarity. ADH: alcohol dehydrogenase, ATA: amino transferase, IRED: imine reductase, CAR: carboxylic acid reductase. R: R-group, limited by enzyme substrate specificity.

#### 5.2.4 Cofactor regeneration

Many oxidoreductases require cofactors, most commonly the comparatively expensive NAD(P)H. For biocatalysts to be economically feasible *in vitro*, cofactor

regeneration is essential. While many enzyme systems have been developed for the regeneration of NAD(P)H, perhaps one of the most useful makes use of the phosphite dehydrogenase (PTDH) enzyme from *Pseudomonas stutzeri*. The wild type enzyme regenerates NADH via the nearly irreversible oxidation of phosphite to phosphate. This enzyme has been engineered to also regenerate NADPH and improve thermostability (Woodyer, Van der Donk and Zhao, 2003; McLachlan, Johannes and Zhao, 2007).

Techniques for the regeneration of ATP are less well developed, but enzymes such as pyruvate kinase (PK), creatine kinase (CK), adenylate kinase (AK) and polyphosphate kinase (PPK) have been used (Zhao and Van Der Donk, 2003). Of these, polyphosphate kinase makes use of the cheapest and most stable substrate, polyphosphate. Polyphosphate strongly chelates  $Mg^{2+}$ , which is an essential cofactor for many enzymes that utilise ATP. This limits the concentration of polyphosphate that can be added to a reaction (Itoh and Shiba, 2004).

Polyphosphate is also a substrate for polyphosphate-AMP phosphotransferase (PAP), which allows the regeneration of ADP from AMP (Woodyer, Johannes and Zhao, 2003). PAP can be combined with an ATP regenerating enzyme for complete regeneration of ATP from AMP. This has demonstrated using both PAP and PPK (Kameda *et al.*, 2001). In an alternative approach, adenylate kinase (which catalyses the reversible conversion of two ADP molecules into one ATP and one AMP) was used in place of PPK. PAP was again used, pushing the equilibrium towards the regeneration of ATP (Resnick and Zehnder, 2000).

### **5.2.5 The application of process modelling**

In developing new biocatalytic processes the use of kinetic modelling is widely advocated, yet often not used in process development (Sin, Woodley and Gernaey, 2009; Ringborg and Woodley, 2016). However, recent literature shows there is increasing interest in this topic (Vasić-Rački, Findrik and Vrsalović Presečki, 2011). The development of a kinetic model early on in the development process can be invaluable in cost / benefit or feasibility analysis (Sin, Woodley and Gernaey, 2009). It can highlight where enzyme engineering could provide substantial gains, allow simulations of hypothetical change that can speed up process development, and allow evidence based decision-making (Rios-Solis *et al.*, 2015). Critically, a model

allows for the identification of bottlenecks and the quantification of process problems (du Preez *et al.*, 2015). For these reasons, the application of validated models is important in driving the field of biocatalytic process development forward (Ringborg and Woodley, 2016). Despite these advantages, few models for multistep biocatalysis processes have been explored. The challenges of integrating kinetic models for multiple enzymes are evident, not least the number of kinetic parameters involved (Rios-Solis *et al.*, 2015).

Two general types of model have been constructed to describe enzyme reactions. “Data driven” empirical methods provide a model of the enzyme reaction without a need to understand the detailed mechanism of the reaction (Park, 2007). Often these models are simpler, involve fewer parameters, and fit the data well, but at the expense of understanding the reaction (Vasić-Rački, Findrik and Vrsalović Presečki, 2011). In multi-enzyme systems where many components are present, protein interactions that could affect the reaction rate may occur. The use of a data driven model can be advantageous in these circumstances (Vasić-Rački, Findrik and Vrsalović Presečki, 2011). Another approach is to use “mechanistic models”. These seek to describe enzyme mechanisms as accurately as possible: a classic example is the use of Michaelis-Menten equations (Peri *et al.*, 2007). As mechanistic models seek to understand a system as well as to predict it, they offer the opportunity to observe unanticipated phenomena, and so generate new hypotheses. These offer opportunities to develop substantial improvements or insight into the development process (Vasić-Rački, Findrik and Vrsalović Presečki, 2011).

Irrespective of whether a pathway is to be applied as a whole-cell system, or as isolated enzymes, there is a need to understand the mechanisms of the enzymes involved and to model reaction kinetics to optimise overall yield (Rios-Solis *et al.*, 2015). Whilst data driven empirical models can be powerful tools for the optimisation of isolated cases, they operate as “black boxes” that offer no further understanding of the system being studied. Where pathways fail to meet the productivity expected of them, mechanistic models can be invaluable in understanding and troubleshooting where a bottleneck may lie (Ringborg and Woodley, 2016).

### 5.2.6 Aims and objectives.

There are many industrially relevant enzymes that generate carboxylic acids or use aldehydes as their substrates (Figure 5-1). The construction of multi-step reactions offers attractive one-pot processes with excellent cost and productivity benefits. To demonstrate the use of CARs in multi-step cascade reactions, and investigate their use *in vitro*, we designed a reaction made up of an esterase, a CAR and an alcohol dehydrogenase (ADH) to hydrolyse and then reduce an ester to its derivative alcohol. Methyl *p*-toluate was chosen as a trial substrate. Enzymes for cofactor regeneration and by-product removal were added resulting in a synthetic seven enzyme system.

To highlight the modular nature of the enzyme toolbox for designing multistep reactions, each enzyme in the system has been characterised individually and a mathematical model for each created. In addition, the effects of pH and temperature were assessed to define an operational window. In constructing and testing the model for each step in the reaction we demonstrate the value of a model in identifying process problems. Specifically, we were able to predict the need for PP<sub>i</sub> removal, and to optimise the trial batch reaction for maximum productivity.

With increasing numbers of enzymes becoming available from commercial sources or described in literature, the development of new cascades will lay the base for artificial metabolisms and create enzyme catalyst networks allowing new reactions to be quickly assembled (Schrittwieser *et al.*, 2011).

## 5.3 Results

### 5.3.1 Expression and purification of enzymes

All of the enzymes used in this study were recombinantly prepared from *E. coli*, as fusions with a polyhistidine-tag. All of the proteins were purified by nickel affinity chromatography followed by size exclusion chromatography. A high level of purity was achieved for each enzyme (Figure 5-2 - Supplementary Figure 1 to Figure 5-2 - Supplementary Figure 8).

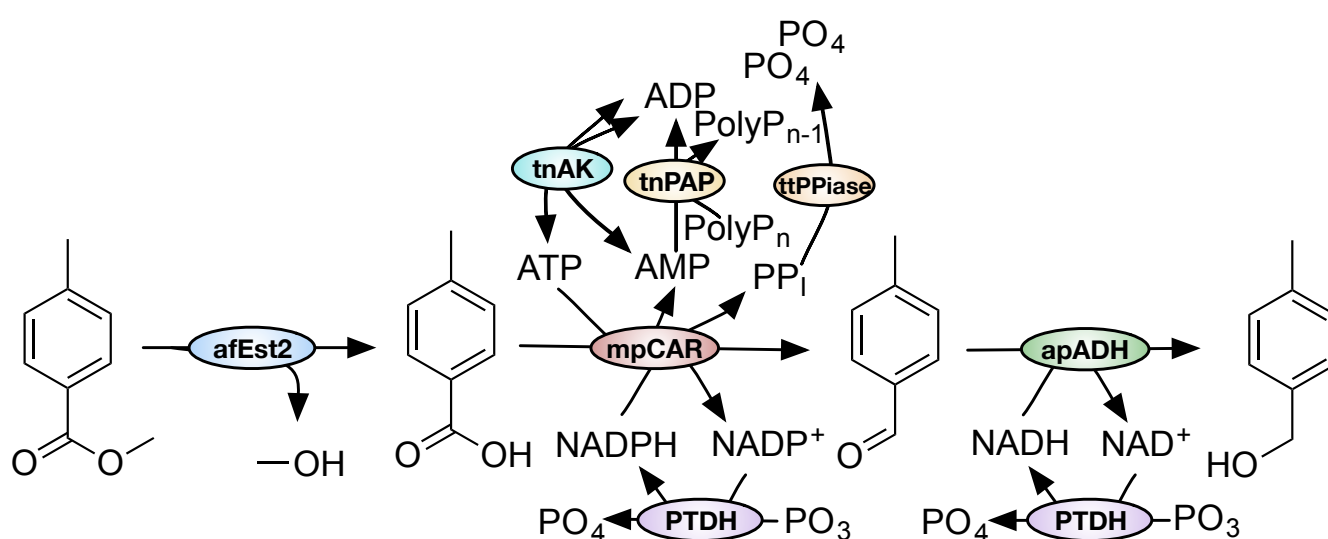
### 5.3.2 Designing a synthetic multi-step pathway

Enzymes with the correct substrate specificity to fill the esterase, CAR and ADH steps were identified from the literature. Where possible, thermostable enzymes were chosen to provide maximum operational stability. In the case of the CAR step, only moderate thermostability was possible due to the limited number of organisms that use these enzymes. An esterase from the hyperthermophile *Archaeoglobus fulgidus* (afEst2), a CAR from the moderate thermophile *Mycobacterium phlei* (mpCAR), and an ADH from the hyperthermophile *Aeropyrum pernix* (apADH) were chosen (Figure 5-2) (Guy, Isupov and Littlechild, 2003; Sayer *et al.*, 2016). The catalytic constants of these enzymes against the relevant substrates in the test pathway were determined (Table 5-2 - Supplementary Figure 1 to Table 5-2 - Supplementary Figure 20), or taken from relevant literature (references shown in Table 5-2).

To allow cofactor regeneration, a thermostable mutant of PTDH, capable of regenerating both NADH and NADPH, was chosen given the need for both of these cofactors in the pathway. For ADP regeneration from AMP, PAP enzymes from thermophilic organisms *Thermodesulfobium narugense* (tnPAP) and *Thermanaerovibrio velox* (tvPAP) were identified with 33 % and 34 % homology against the previously characterised PAP from *Acinetobacter johnsonii* (Bonting, Kortstee and Zehnder, 1991). Both showed PAP activity (Figure 5-2 - Supplementary Figure 9). tnPAP was used in the final pathway due to greater activity.



A thermostable PPT enzyme from *Thermosynechococcus elongatus* that had previously been characterised (Sato *et al.*, 2007) was initially chosen to regenerate ATP from ADP. However, in our hands this enzyme gave very low activity. In its place, a thermostable AK enzyme from *Thermotoga neapolitana* (tnAK) was used, which had previously been characterised with high activity (Vieille *et al.*, 2003). Upon identifying a need for the removal of PP<sub>i</sub> from the reaction, a thermostable PPiase from *Thermus thermophilus* was also added (ttPPiase) and its activity determined (Table 5-2 - Supplementary Figure 15).

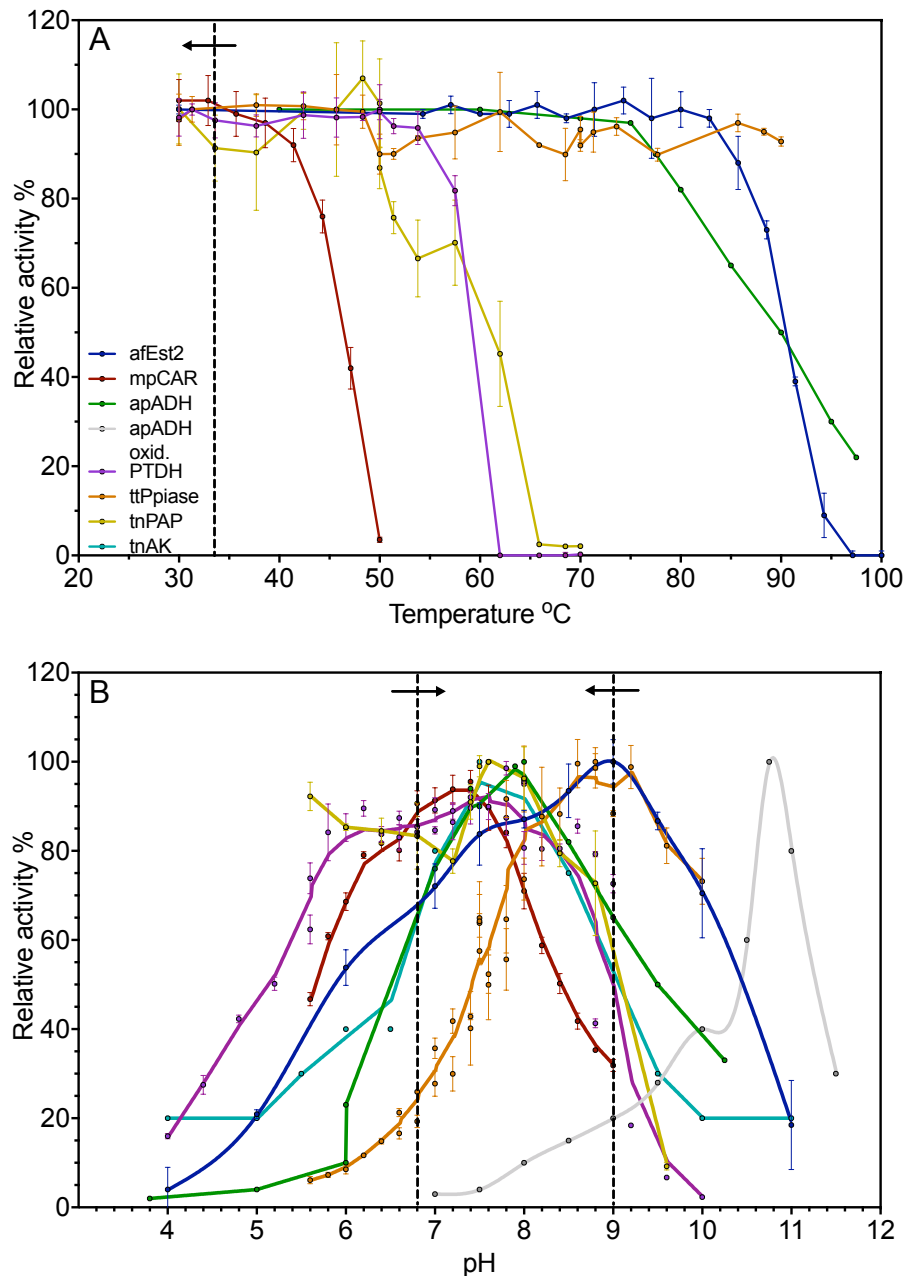


**Figure 5-2 - A schematic of the seven enzyme reaction**

The hydrolysis of methyl *p*-toluate to *p*-toluic acid, followed by reduction to *p*-tolualdehyde and further to *p*-tolylalcohol is shown. The use or production of water is not shown. afEst2: Esterase enzyme from *Archaeoglobus fulgidus*, mpCAR: Carboxylic acid reductase from *Mycobacterium phlei*, apADH: Alcohol dehydrogenase from *Aeropyrum pernix*, PTDH: Engineered phosphite dehydrogenase from *Pseudomonas stutzeri*, ttPPiase: Inorganic pyrophosphatase from *Thermus thermophilus*, tnPAP: Polyphosphate AMP phosphotransferase from *Thermodesulfobium narugense*, tnAK: Adenylate kinase from *Thermotoga neapolitana*. PolyP<sub>n</sub>: A polyphosphate molecule of with a chain length of *n* phosphates.

### 5.3.3 Identifying an operational window.

Thermostability and activity at pH was determined for each enzyme to define an operational window for the reaction (Figure 5-3). Data for afEst2(Sayer *et al.*, 2016), mpCAR(Finnigan *et al.*, 2016), apADH(Hirakawa, Kamiya and Kowarabayashi, 2004) and tnAK(Vieille *et al.*, 2003) was adapted from previous work. The effects of pH and temperature on the remaining enzymes were characterised (Figure 5-3 - Supplementary Figure 1 to Figure 5-3 - Supplementary Figure 6). Taking a threshold of 25 % relative activity or more for each enzyme, a window between pH 6.8 and pH 9.0 was identified. pH 7.5 was chosen as a compromise between all enzymes, and as ADH activity in the (undesired) oxidative direction was minimised. As we expected, the operational window for temperature is primarily limited by the mpCAR enzyme, as this has only moderate thermostability. A reaction temperature of 30 °C was chosen to ensure maximum activity of this enzyme.



**Figure 5-3 - The operational window for temperature (A) and pH (B) for the seven enzyme reaction**

- A. Residual relative activity after incubation at various temperatures for 30 minutes. The operational window for temperature is shown by the dashed vertical line and arrow.
- B. Relative activity at various pH values. Values are relative to the maximum activity in each case. The operational window is shown by two dashed vertical lines, and arrows pointing inwards.

### 5.3.4 Reaction modelling.

Mathematical models utilising kinetics based on the steady state approximation were developed in isolation for each section of the multi-step reaction, and validated before being combined into a multi-step process. These models were developed using Python scripts (Supplementary File 1) following a single approach, with equations appropriate for each enzyme. The rate equations are summarised in Table 5-1, which were used to construct a full set of differential equations (Table 1 - Supplementary Figure 1). Kinetic parameters were either identified from available literature or calculated experimentally. An uncertainty analysis was carried out for each model, where possible using bounds equal to the 95 % confidence intervals of each parameter (Table 5-2). In some cases where uncertainty was judged to be large due to incomplete knowledge, bounds of  $\pm 50$  % of the parameter were used. Starting concentrations were given bounds of  $\pm 5$  %, except in the case of polyphosphate where the length of the polyphosphate chain is unknown. In this case  $\pm 25$  % was used with the upper bound equal to the absolute concentration of phosphate units. For more information on the use of uncertainty and sensitivity analysis in process modelling please see the following references (Sin, Gernaey and Lantz, 2009; Price *et al.*, 2013). Model predictions were tested by running small scale reactions in a thermomixer, with samples taken every 30 minutes into acetonitrile for analysis by HPLC.

### Table 5-1 – Kinetic equations

The rationale for the kinetic equations chosen for each enzyme and interacting partner are discussed in detail below. Aldehyde degradation was modelled as a first order process.

#### afEst2

$$r1 = c_{Est} \cdot k_{cat}^{Est} \cdot \frac{c_{Ester}}{c_{Ester} + K_M^{Ester}}$$

#### mpCAR

$$K_M^{ATP} = K_M^{ATP} \cdot \frac{1 + \frac{c_{PPI}}{K_I^{PPI-Acid}}}{1 + \frac{c_{PPI}}{\alpha PPI \cdot K_I^{PPI-Acid}}} \cdot \left(1 + \frac{c_{AMP}}{K_I^{AMP}}\right) \cdot \left(1 + \frac{c_{ADP}}{K_I^{ADP}}\right)$$

$$K_M^{NADPH} = K_M^{NADPH} \cdot \left(1 + \frac{c_{NADP+}}{K_I^{NADP+}}\right)$$

$$K_M^{Acid} = K_M^{Acid} \cdot \left(1 + \frac{c_{PPI}}{K_I^{PPI-Acid}}\right)$$

$$k_{cat}^{CAR} = \frac{k_{cat}^{CAR}}{1 + \frac{c_{PPI}}{\alpha PPI \cdot K_I^{PPI-Acid}}}$$

$$r2 = c_{CAR} \cdot k_{cat}^{CAR} \cdot \frac{c_{Acid} \cdot c_{ATP}}{(K_I^{ATP} \cdot K_M^{Acid}) + (K_M^{Acid} \cdot c_{ATP}) + (K_M^{ATP} \cdot c_{Acid}) + (c_{ATP} \cdot c_{Acid})} \cdot \frac{c_{NADPH}}{c_{NADPH} + K_m^{NADPH}}$$

#### apADH

$$r3 = c_{ADH} \cdot k_{cat}^{ADHFwd} \cdot \frac{c_{Aldehyde} \cdot c_{NADH}}{(K_I^{NADH} \cdot K_M^{Aldehyde}) + (K_M^{Aldehyde} \cdot c_{NADH}) + (K_M^{NADH} \cdot c_{Aldehyde}) + (c_{NADH} \cdot c_{Aldehyde})}$$

$$r4 = c_{ADH} \cdot k_{cat}^{ADHRev} \cdot \frac{c_{Alcohol} \cdot c_{NAD+}}{(K_I^{NAD+} \cdot K_M^{Alcohol}) + (K_M^{Alcohol} \cdot c_{NAD+}) + (K_M^{NAD+} \cdot c_{Alcohol}) + (c_{NAD+} \cdot c_{Alcohol})}$$

#### PTDH

$$K_M^{NAD+} = K_M^{NAD+} \cdot \left(1 + \frac{c_{NADP+}}{K_I^{NADP+}}\right)$$

$$K_M^{NADP+} = K_M^{NADP+} \cdot \left(1 + \frac{c_{NAD+}}{K_I^{NAD+}}\right)$$

$$r5 = c_{PTDH} \cdot k_{cat}^{PTDH-NAD+} \cdot \frac{c_{NAD+}}{c_{NAD+} + K_M^{NAD+}}$$

$$r6 = c_{PTDH} \cdot k_{cat}^{PTDH-NADP+} \cdot \frac{c_{NADP+}}{c_{NADP+} + K_M^{NADP+}}$$

#### PPiase

$$r7 = c_{PPiase} \cdot k_{cat}^{PPiase} \cdot \frac{c_{PPI}}{c_{PPI} + K_M^{PPI}}$$

#### PAP

$$r8 = c_{PPT} \cdot k_{cat}^{PPT} \cdot \frac{c_{PolyP}}{c_{PolyP} + K_M^{PolyP}} \cdot \frac{c_{AMP}}{c_{AMP} + K_M^{AMP}}$$

#### AK

$$r9 = c_{AK} \cdot k_{cat}^{AK-Fwd} \cdot \frac{c_{ADP}}{c_{ADP} + K_M^{ADP}} \cdot \frac{c_{ADP}}{c_{ADP} + K_M^{ADP}}$$

$$r10 = c_{AK} \cdot k_{cat}^{AK-Rev} \cdot \frac{c_{AMP}}{c_{AMP} + K_M^{AMP}} \cdot \frac{c_{ATP}}{c_{ATP} + K_M^{ATP}}$$

#### Aldehyde degradation

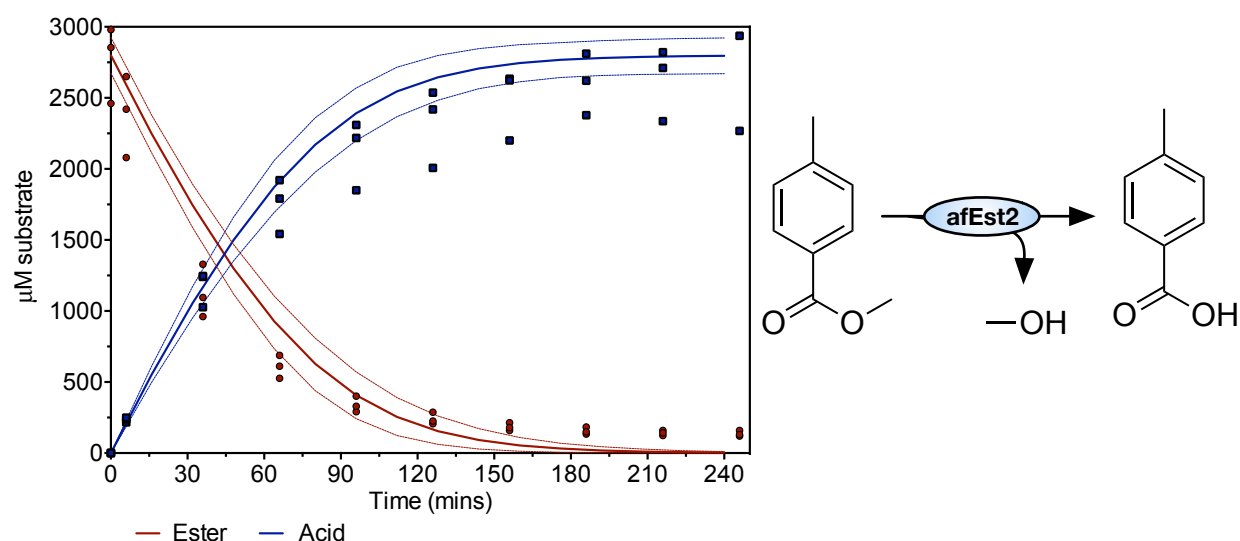
$$r11 = k^{ad} \cdot c_{Aldehyde}$$

Table 5-2 – Kinetic parameters used in the model.	
<b>afEst2</b> (Sayer <i>et al.</i> , 2016)	
$k_{cat}$ (Sayer <i>et al.</i> , 2016)	$6 \pm 1 \text{ min}^{-1}$
$K_{M-Ester}$ (Sayer <i>et al.</i> , 2016)	$1,500 \pm 200 \text{ }\mu\text{M}$
<b>mpCAR</b> (Finnigan <i>et al.</i> , 2016)	
$k_{cat}$	$200 \pm 20 \text{ min}^{-1}$
$K_{M-Acid}$	$1,500 \pm 320 \text{ }\mu\text{M}$
$K_{M-ATP}$	$100 \pm 28 \text{ }\mu\text{M}$
$K_{I-ATP}$	$40 \pm 34 \text{ }\mu\text{M}$
$K_{M-NADPH}$	$30 \pm 8$
$K_{I-AMP}$ (Finnigan <i>et al.</i> , 2016)	$10,000 \pm 1,800$
$K_{I-ADP}$	$11,000 \pm 4,000$
$K_{I-NADP+}$ (Finnigan <i>et al.</i> , 2016)	$143 \pm 16$
$K_{I-PPi-Acid}$ (Finnigan <i>et al.</i> , 2016)	$340 \pm 80 \text{ }\mu\text{M}$
$K_{I-PPi-ATP}$ (Finnigan <i>et al.</i> , 2016)	$220 \pm 100 \text{ }\mu\text{M}$
$\alpha_{-PPi-ATP}$ (Finnigan <i>et al.</i> , 2016)	$2.6 \pm 2.8$
<b>apADH</b>	
$k_{cat-Fwd}$	$1.7 \pm 0.2 \text{ min}^{-1}$
$k_{cat-Rev}$	$1.7 \pm 0.85 \text{ min}^{-1}$
$K_{M-NADH}$	$180 \pm 60 \text{ }\mu\text{M}$
$K_{M-NAD+}$	$190 \pm 40 \text{ }\mu\text{M}$
$K_{I-NAD+}$	$185 \pm 92.5 \text{ }\mu\text{M}$
$K_{I-NADH}$	$185 \pm 92.5 \text{ }\mu\text{M}$
$K_{M-Aldehyde}$	$350 \pm 120 \text{ }\mu\text{M}$
$K_{M-Alcohol}$	$10,000 \pm 5,000$
<b>ttPpiase</b> (Schomburg <i>et al.</i> , 2013)	
$k_{cat}$	$4,400 \pm 2,200 \text{ min}^{-1}$
$K_{M-PPi}$ (Schomburg <i>et al.</i> , 2013)	$500 \pm 250 \text{ }\mu\text{M}$
<b>PTDH</b>	
$k_{cat-NAD+}$	$637 \pm 16 \text{ min}^{-1}$
$k_{cat-NADP+}$	$342 \pm 16 \text{ min}^{-1}$
$K_{M-NAD+}$	$85 \pm 5 \text{ }\mu\text{M}$
$K_{M-NADP+}$	$220 \pm 40 \text{ }\mu\text{M}$
<b>tnPAP</b>	

$k_{cat}$	$250 \pm 60 \text{ min}^{-1}$
$K_{M-AMP}$	$280 \pm 100 \text{ }\mu\text{M}$
$K_{M-PolyP}$	$4,000 \pm 1,400 \text{ }\mu\text{M}$
<b>tnAK</b> (Perrier <i>et al.</i> , 1998; Vieille <i>et al.</i> , 2003)	
$k_{cat-ADP}$ (Vieille <i>et al.</i> , 2003)	$2,340 \pm 1,170 \text{ min}^{-1}$
$k_{cat-AMP-ATP}$ (Vieille <i>et al.</i> , 2003)	$3,950 \pm 1,975 \text{ min}^{-1}$
$K_{M-ADP}$ (Perrier <i>et al.</i> , 1998)	$91 \pm 45.5 \text{ }\mu\text{M}$
$K_{M-AMP}$ (Perrier <i>et al.</i> , 1998)	$38 \pm 19 \text{ }\mu\text{M}$
$K_{M-ATP}$ (Perrier <i>et al.</i> , 1998)	$51 \pm 25.5 \text{ }\mu\text{M}$
<b>Aldehyde degradation</b>	
$k$	$0.00279 \pm 0.001395 \text{ min}^{-1}$
Parameters were determined experimentally or obtained from the literature. Errors represent 95% confidence intervals where these could be experimentally determined. Where these could not be experimentally determined, errors of 50 % of the parameter value were conservatively assigned.	

### 5.3.5 Esterase reaction

afEst2 was recently characterised and kinetic parameters for the hydrolysis of methyl *p*-toluate at 30 °C reported (Sayer *et al.*, 2016) (Table 5-2). The hydrolysis reaction is described by the one substrate Michaelis-Menten equation (Table 5-1). The model was tested against a small scale batch reaction (Figure 5-4), which it was able to predict well. It was noted that the reaction proceeds only to approximately 95% completion, where there is a plateau. This was seen for several other reactions.



**Figure 5-4 - Validation of the esterase model.**

A reaction containing 10  $\mu\text{M}$  afEst2 and 2,800  $\mu\text{M}$  methyl *p*-toluate (ester) was used to validate the model for afEst2. Ester and acid concentrations, measured every 30 minutes by HPLC, are shown as red circles and blue squares respectively. The model prediction is shown as the solid line in the same colours. Dashed lines represent the 5<sup>th</sup> and 95<sup>th</sup> percentile of the uncertainty analysis.

### 5.3.6 CAR reaction

The CAR enzymes have three substrates: ATP, NADPH and a carboxylic acid. mpCAR has a low  $K_M$  for NADPH (Finnigan *et al.*, 2016) and it is expected that in the coupled reactions, the NADPH concentration will considerably exceed this at all times. This reaction was therefore modelled considering only the ATP and carboxylic



acid concentrations that are relevant for the rate limiting step (Finnigan *et al.*, 2016). A steady state sequential ordered bi-substrate equation was used to model ATP and carboxylic acid concentration effects. The ordered binding of ATP followed by a carboxylic acid has been shown previously for CAR enzymes (Finnigan *et al.*, 2016), and is confirmed by the data in this study. This was combined with a term equating to the Michaelis-Menten equation to model the effects of NADPH concentration. Parameters were determined experimentally (Table 5-2 - Supplementary Figure 3 and Table 5-2 - Supplementary Figure 4).

A reaction to test the CAR model was set up, predicting a fast turnover of the acid into its derivative aldehyde (shown in grey, Figure 5-5A). Instead, after an initial burst, activity slowed considerably and the reaction was not complete even after four hours. We investigated the possibility of product inhibition and found that all of the products of the CAR reaction ( $PP_i$ , AMP and  $NADP^+$ ) act as inhibitors (discussed in more detail in (Finnigan *et al.*, 2016)). While AMP and  $NADP^+$  both act as competitive inhibitors which may be overcome by increasing the substrate concentration,  $PP_i$  acts as a mixed model inhibitor with respect to ATP, causing a large decrease in the apparent  $k_{cat}$ . When the inhibition by these products was taken into account the model fitted the data quite well (Figure 5-5A). We noted that modelling the inhibition resulted in a large level of uncertainty.

### 5.3.7 ttPPIase removes CAR inhibition by $PP_i$

In order to alleviate inhibition by  $PP_i$  on the CAR enzyme, an inorganic pyrophosphatase from *Thermus thermophilus* (ttPPIase) was added to the reaction. This enzyme shows exceptional thermostability and retained over 90 % activity after a 30 minute incubation at 95 °C (Figure 5-3). Pyrophosphatase activity was measured at a saturating concentration of  $PP_i$  (5 mM) and the rate taken to be the  $k_{cat}$ . The enzyme was modelled using the one substrate Michaelis-Menten equation, and using the  $K_M$  determined in a recent work on this enzyme (Mu *et al.*, 2009). The addition of the ttPPIase alleviated the majority of the inhibitory effects seen previously and the CAR reaction went to completion within one hour (Figure 5-5B).

### 5.3.8 Aldehyde degradation

Once the CAR reaction was complete, aldehyde concentration decreased steadily over time (Figure 5-5B). The product of this degradation could not be identified by HPLC, indicating that it was either lost from the reaction solution, or was not detectable in this HPLC mode. A one phase decay equation was fitted to the decrease in aldehyde concentration, and an equation for the rate of aldehyde degradation constructed (Table 5-1 and Table 5-2). This resulted in a very good fit of the aldehyde concentration to the data.

### 5.3.9 NAD(P)H regeneration using PTDH

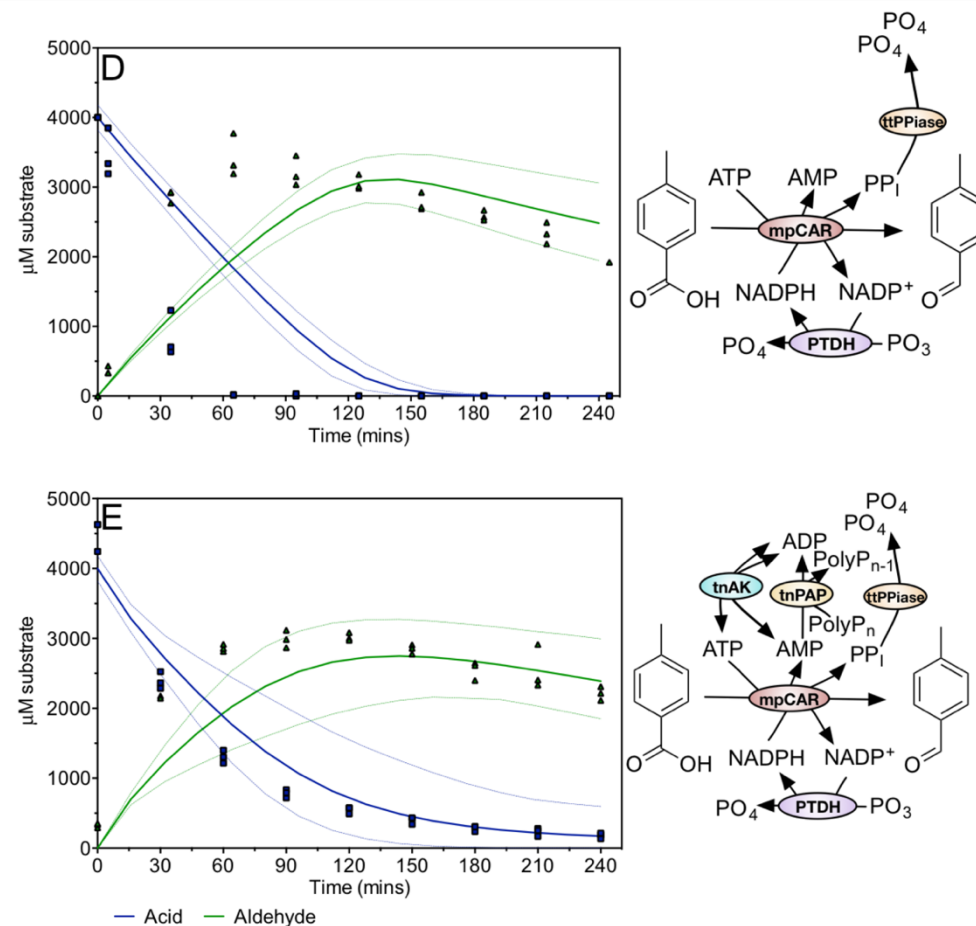
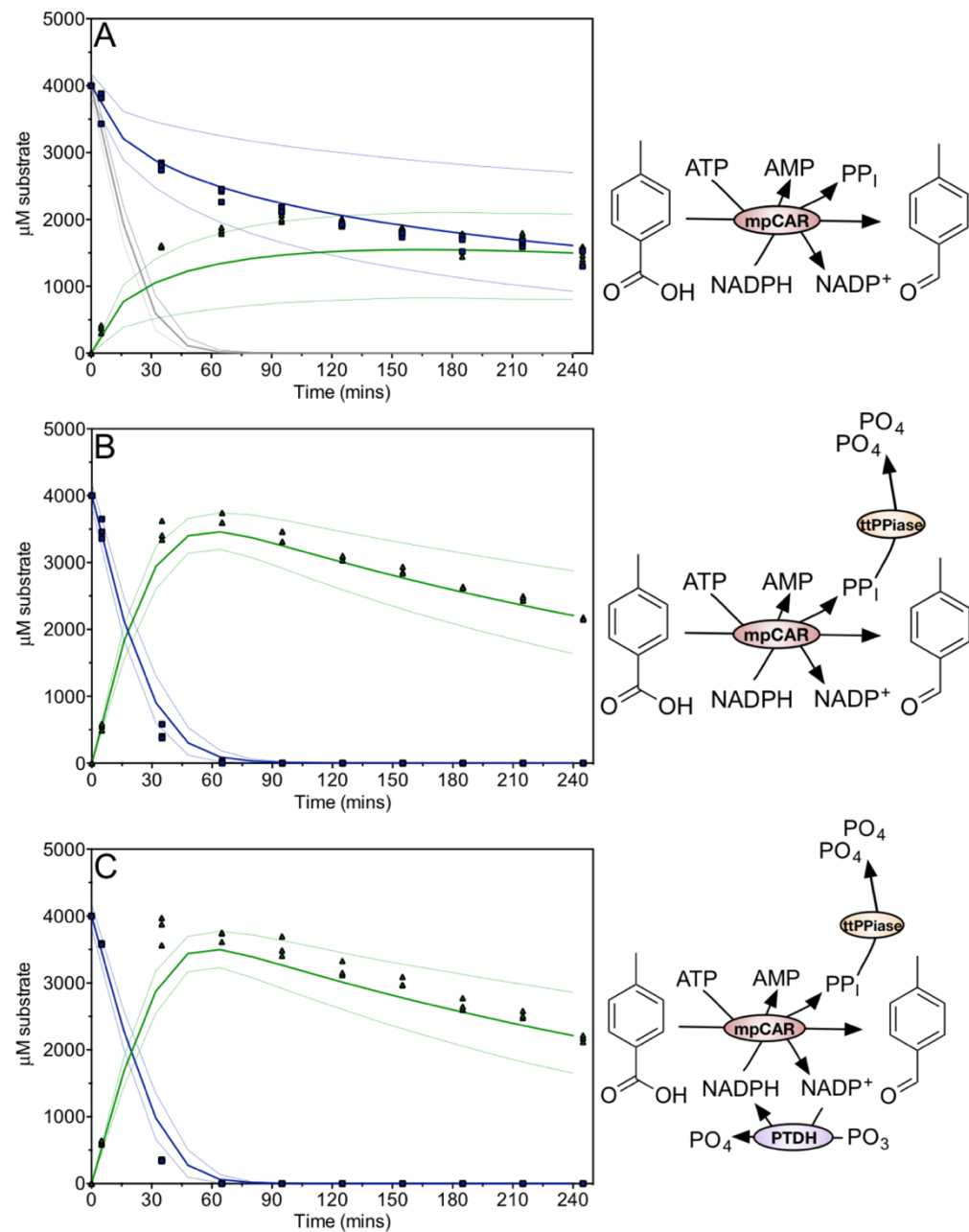
The regeneration of NADPH using PTDH was then added to the CAR step. Although PTDH shows sequential ordered kinetics, with  $\text{NAD}^+$  binding before phosphite, suitable parameters to use a two substrate equation are not reported in the literature. The phosphite concentration was not modelled, as all experiments will be performed at saturating phosphite concentrations throughout. The one substrate Michaelis-Menten equation was therefore used for reactions with either  $\text{NAD}^+$  or  $\text{NADP}^+$  as the substrate.  $\text{NAD}^+$  and  $\text{NADP}^+$  act as competitive inhibitors of each other with  $K_i$  values equal to their respective  $K_M$  values, and this inhibition was added to the model (Table 5-1). Kinetic parameters for  $\text{NAD}^+$ ,  $\text{NADP}^+$  were determined at assay conditions by following the generation of NADH or NADPH at 340 nm (Table 5-2; Table 5-2 - Supplementary Figure 13 Table 5-2 - Supplementary Figure 14). The  $K_M$  for  $\text{NAD}^+$  was in good agreement with previously reported data for the thermostable mutant PTDH. However our enzyme showed a substantially higher  $k_{cat}$  for  $\text{NAD}^+$  than previously reported (McLachlan, Johannes and Zhao, 2007).

The NADPH regeneration was tested by attempting reduction of 4000  $\mu\text{M}$  *p*-toluic acid by CAR with only 500  $\mu\text{M}$  or 50  $\mu\text{M}$  NADPH, together with the PTDH regeneration system. Both reactions went to completion and the reaction containing 500  $\mu\text{M}$  NADPH fit model predictions well (Figure 5-5C). When only 50  $\mu\text{M}$  NADPH was used, a much slower reaction was predicted by the model. However, only a small difference was observed experimentally compared to the use of 500  $\mu\text{M}$  NADPH (Figure 5-5D). We considered that some form of substrate channelling (Lin, Palomec and Wheeldon, 2014; Wheeldon *et al.*, 2016) may be taking place between the CAR and PTDH resulting in a lower apparent  $K_M$  for  $\text{NADP}^+$  by PTDH.

### 5.3.10 ATP regeneration using tnPAP and tnAK.

Both tvPAP and tnPAP showed good PAP activity (Figure 5-2 - Supplementary Figure 9). tnPAP showed the best activity and so was chosen for further characterisation. Kinetic parameters for polyphosphate and AMP were determined experimentally (Table 5-2 - Supplementary Figure 17, Table 5-2 - Supplementary Figure 18, Table 5-2 - Supplementary Figure 19), and a bi-substrate equation constructed.

An adenylate kinase from *Thermotoga neapolitana* has been reported to possess excellent thermostability, with kinetic parameters similar to those of the characterised *E. coli* adenylate kinase. These parameters were used to model the adenylate kinase. However, testing of this enzyme in a CAR coupled reaction revealed a significantly lower turnover number (Table 5-2 - Supplementary Figure 21). Furthermore, magnesium concentration has been shown to be instrumental in controlling the equilibrium catalysed by AK. Since the free magnesium concentration is not known due to chelation by polyphosphate, this effect could not be taken into account. For this reason, the AK reaction is difficult to model accurately in the context of the multistep pathway. Therefore, a simple bi-substrate equation was used to describe it. Large uncertainty bounds were used for all the parameters in the AK reaction as a consequence.



### Figure 5-5 - Validation of mpCAR and related enzyme models.

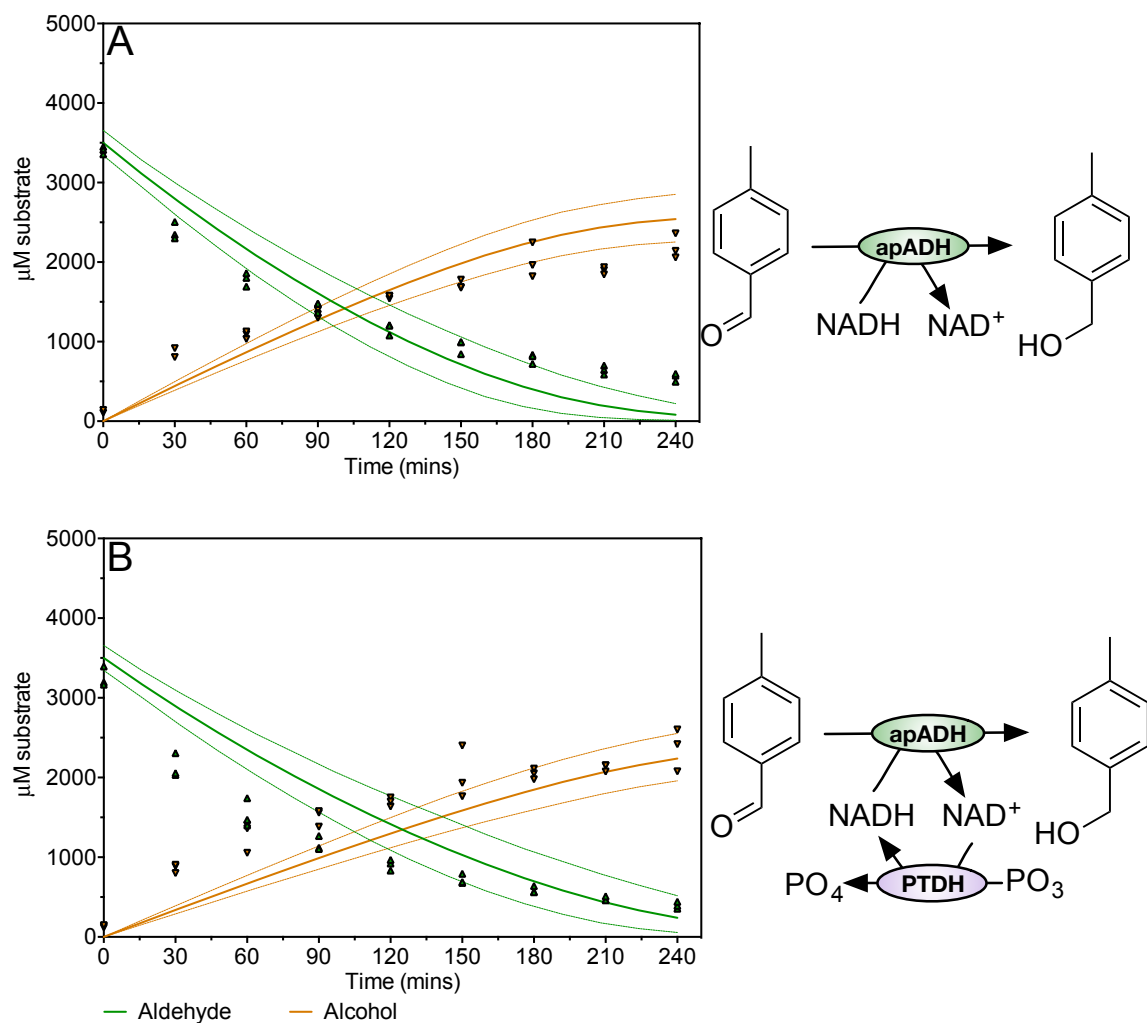
Reactions were performed to validate the model for mpCAR, and the addition of cofactor regenerating and PP<sub>i</sub> removing enzymes. Acid and aldehyde concentrations measured every 30 minutes by HPLC are shown as blue squares and green upwards facing triangles respectively. The model prediction is shown as the solid line in the same colours. Dashed lines represent the 5<sup>th</sup> and 95<sup>th</sup> percentile of the uncertainty analysis.

- A. mpCAR alone. Reactions were initiated with 1  $\mu\text{M}$  mpCAR, 8,000  $\mu\text{M}$  ATP, 5,000  $\mu\text{M}$  NADPH, 20,000  $\mu\text{M}$  MgCl<sub>2</sub>, 4,000  $\mu\text{M}$  *p*-toluic acid. The initial model prediction, which did not take in to account any product inhibition, is shown in grey.
- B. mpCAR-ttPpiase. Reactions were initiated with 1  $\mu\text{M}$  mpCAR, 1  $\mu\text{M}$  ttPpiase, 8,000  $\mu\text{M}$  ATP, 5,000  $\mu\text{M}$  NADPH, 20,000  $\mu\text{M}$  MgCl<sub>2</sub>, 4,000  $\mu\text{M}$  *p*-toluic acid.
- C. mpCAR-ttPpiase-PTDH. Reactions were initiated with 1  $\mu\text{M}$  mpCAR, 1  $\mu\text{M}$  ttPpiase, 1  $\mu\text{M}$  PTDH, 8,000  $\mu\text{M}$  ATP, 500  $\mu\text{M}$  NADPH, 20,000  $\mu\text{M}$  MgCl<sub>2</sub>, 20,000  $\mu\text{M}$  PO<sub>3</sub>, 4,000  $\mu\text{M}$  *p*-toluic acid.
- D. mpCAR-ttPpiase-PTDH (low NADPH). Reactions were initiated with 1  $\mu\text{M}$  mpCAR, 1  $\mu\text{M}$  ttPpiase, 1  $\mu\text{M}$  PTDH, 8,000  $\mu\text{M}$  ATP, 50  $\mu\text{M}$  NADPH, 20,000  $\mu\text{M}$  MgCl<sub>2</sub>, 20,000  $\mu\text{M}$  PO<sub>3</sub>, 4,000  $\mu\text{M}$  *p*-toluic acid.
- E. mpCAR-ttPpiase-PTDH-tnPAP-tnAK. Reactions were initiated with 0.4  $\mu\text{M}$  mpCAR, 1  $\mu\text{M}$  ttPpiase, 1  $\mu\text{M}$  PTDH, 3  $\mu\text{M}$  tnPAP, 1  $\mu\text{M}$  tnAK, 1,250  $\mu\text{M}$  ATP, 500  $\mu\text{M}$  NADPH, 20,000  $\mu\text{M}$  MgCl<sub>2</sub>, 20,000  $\mu\text{M}$  PO<sub>3</sub>, 6,000  $\mu\text{M}$  polyphosphate, 4,000  $\mu\text{M}$  *p*-toluic acid.

### 5.3.11 ADH reaction

Kinetic parameters for both the forward and reverse directions of the ApADH catalysed reactions were determined experimentally. The  $K_M$  for *p*-tolyl alcohol (for the reverse oxidative reaction) could not be accurately determined, as it is likely outside the range which can be tested (Table 5-2 - Supplementary Figure 11 and Table 5-2 - Supplementary Figure 12). An estimated  $K_M$  of 100 mM was used (Table 5-2). Consequently,  $k_{cat-Rev}$  was also difficult to determine and a value similar to  $k_{cat-Fwd}$  was assumed. A two substrate steady state sequential rate equation was used to describe both the forward and reverse reactions separately, using an estimated parameter for the  $K_I$  for  $NAD^+$  and NADH based on their determined  $K_M$  values.

Reactions featuring only apADH as well as with PTDH for NADH regeneration were used to validate the model, which predicted the rate of reduction well in both cases (Figure 5-6). A sensitivity analysis of the final alcohol concentration in both these reactions showed the  $K_I$  for  $NAD^+$  and NADH, and  $k_{cat-Rev}$  to have little effect (Figure 5-6 - Supplementary Figure 1).



**Figure 5-6 - Validation of ADH and ADH-PTDH modals**

Reactions used to validate the model for apADH, and cofactor regeneration by PTDH. Aldehyde and alcohol concentrations measured every 30 minutes are shown as green upwards facing triangles and orange downward facing triangles respectively. The model prediction is shown as the solid line in the same colours. Dashed lines represent the 5<sup>th</sup> and 95<sup>th</sup> percentile of the uncertainty analysis.

apADH only. Reactions were initiated with 10  $\mu\text{M}$  apADH, 5,000  $\mu\text{M}$  NADH, 3,500  $\mu\text{M}$  *p*-tolualdehyde.

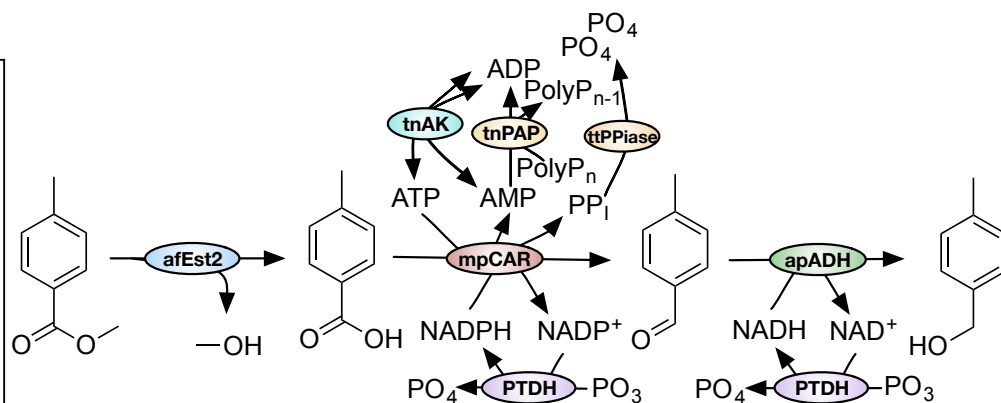
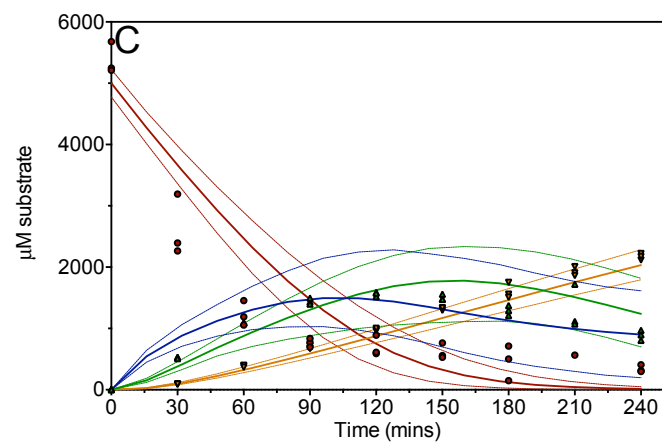
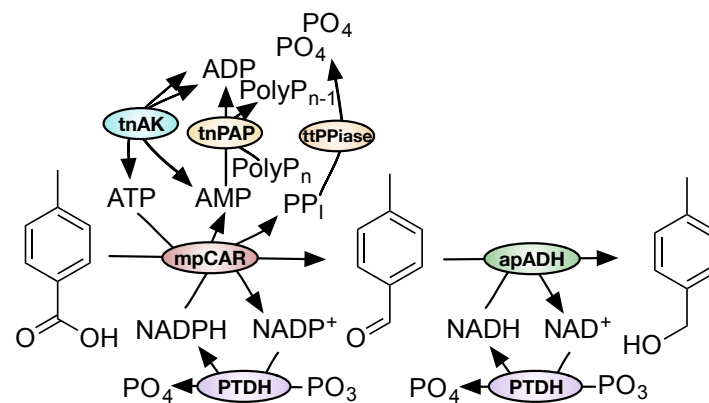
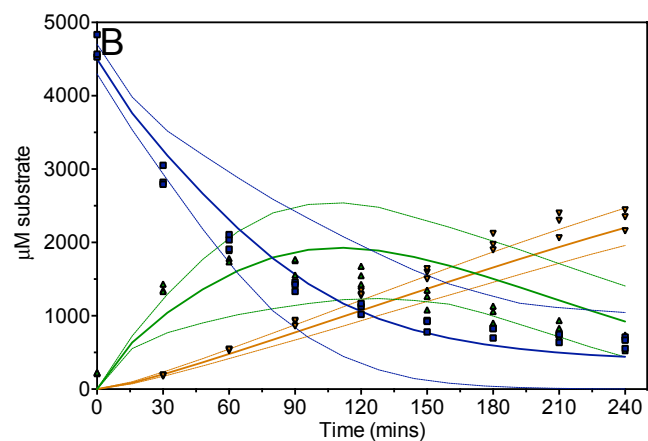
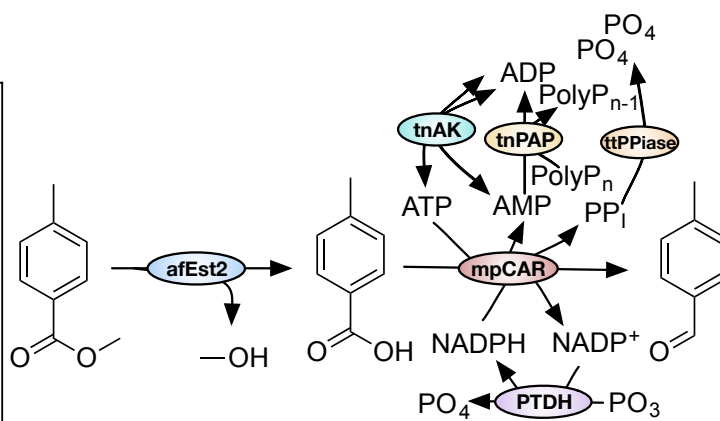
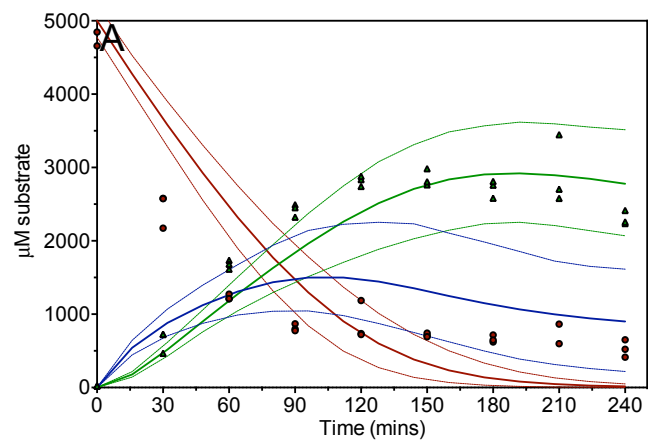
apADH-PTDH. Reactions were initiated with 10  $\mu\text{M}$  apADH, 1  $\mu\text{M}$  PTDH, 20,000  $\mu\text{M}$   $\text{PO}_3$ , 500  $\mu\text{M}$  NADH, 3,500  $\mu\text{M}$  *p*-tolualdehyde.

### **5.3.12 Building a multi-step enzyme cascade**

Once the models for each of the esterase, CAR and ADH steps were independently validated with cofactor regeneration, multi-step reactions were constructed. The CAR step was tested in combination with first the esterase, and then ADH (Figure 5-7A, B). In each case the model performed well in predicting the productivity of the reaction, although the hydrolysis reaction catalysed by afEst2 proceeded faster than was expected (Figure 5-7A). This might again reflect a substrate channelling effect speeding the afEst2 reaction when CAR is available to receive the product, or the activation of afEst2 by one of the substrates not included when afEst2 was tested alone. The full desired multistep reaction was then constructed (Figure 5-7C). Again, the afEst2 reaction appeared to be proceeding at approximately twice the expected rate. However the data showed an excellent fit in terms of the final product measurements.







— Ester — Acid — Aldehyde — Alcohol

### Figure 5-7 - Validations of combinations of the esterase, CAR and ADH reactions.

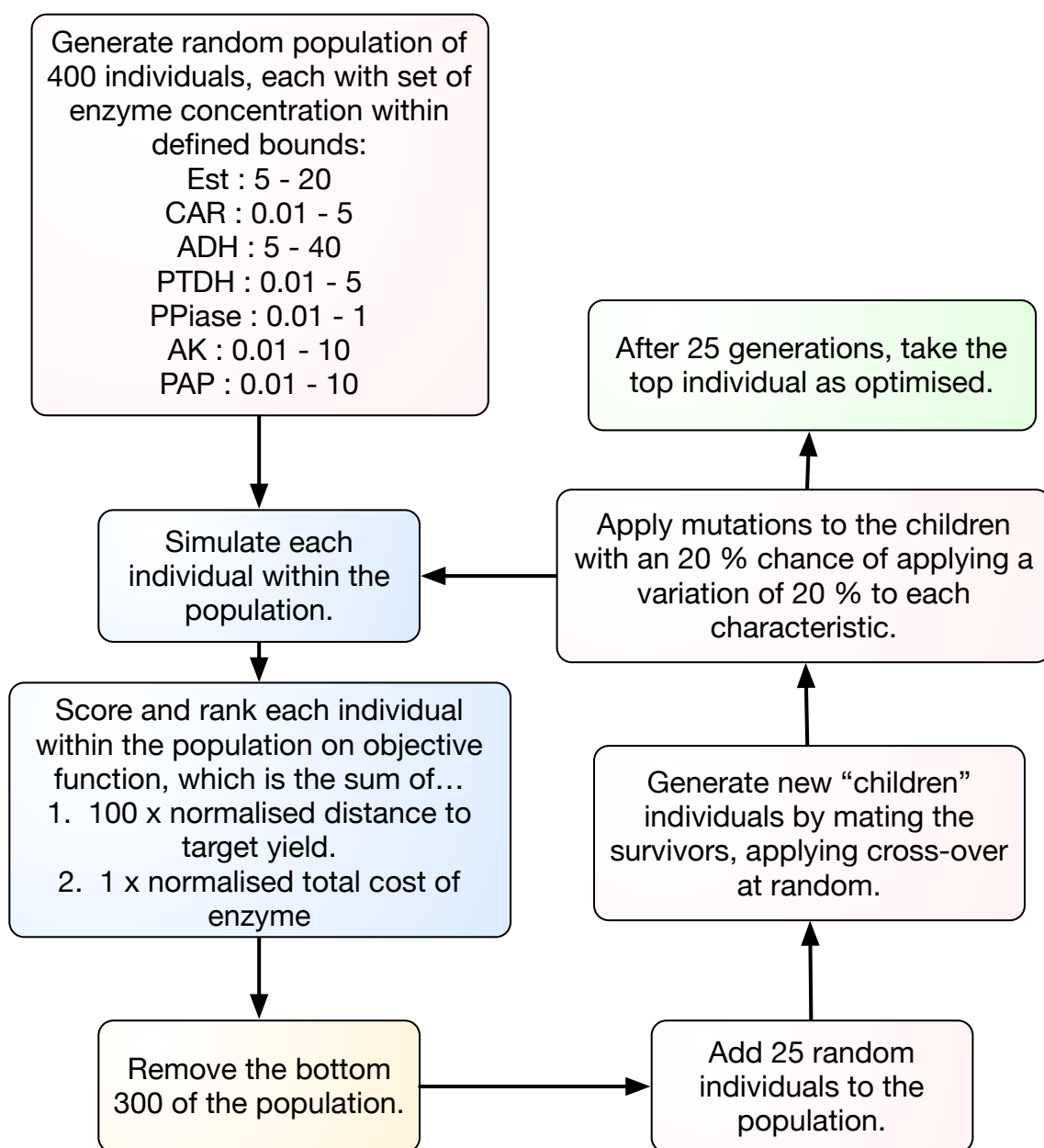
Reactions combining the CAR step with the esterase, ADH or both were constructed to test the model for these multi-step reactions. Ester, acid, aldehyde and alcohol concentrations measured every 30 minutes are shown as red circles, blue squares, green upwards facing triangles and orange downward facing triangles respectively. The model prediction is shown as the solid line in the same colours. Dashed lines represent the 5<sup>th</sup> and 95<sup>th</sup> percentile of the uncertainty analysis.

- A. afEst2-mpCAR-ttPpiase-PTDH-tnPAP-tnAK. Reactions were initiated with 10  $\mu$ M afEst2, 0.4  $\mu$ M mpCAR, 1  $\mu$ M ttPpiase, 1  $\mu$ M PTDH, 3  $\mu$ M tnPAP, 1  $\mu$ M tnAK, 5,000  $\mu$ M methyl *p*-toluate, 500  $\mu$ M NADPH, 1,250  $\mu$ M ATP, 20,000 phosphite, 6,000  $\mu$ M polyphosphate, 20,000 MgCl<sub>2</sub>.
- B. mpCAR-ttPpiase-PTDH-tnPAP-tnAK-apADH. Reactions were initiated with 0.4  $\mu$ M mpCAR, 1  $\mu$ M ttPpiase, 1  $\mu$ M PTDH, 3  $\mu$ M tnPAP, 1  $\mu$ M tnAK, 10  $\mu$ M apADH, 4,500  $\mu$ M *p*-toluic acid, 500  $\mu$ M NADPH, 500  $\mu$ M NADH, 1,250  $\mu$ M ATP, 20,000 phosphite, 6,000  $\mu$ M polyphosphate, 20,000 MgCl<sub>2</sub>.
- C. afEst2-mpCAR-ttPpiase-PTDH-tnPAP-tnAK-apADH. Reactions were initiated with 10  $\mu$ M afEst2, 0.4  $\mu$ M mpCAR, 1  $\mu$ M ttPpiase, 1  $\mu$ M PTDH, 3  $\mu$ M tnPAP, 1  $\mu$ M tnAK, 10  $\mu$ M apADH, 5,000  $\mu$ M methyl *p*-toluate, 500  $\mu$ M NADPH, 500  $\mu$ M NADH, 1,250  $\mu$ M ATP, 20,000 phosphite, 6,000  $\mu$ M polyphosphate, 20,000 MgCl<sub>2</sub>.

### 5.3.14 Optimisation of enzyme concentrations by a genetic algorithm

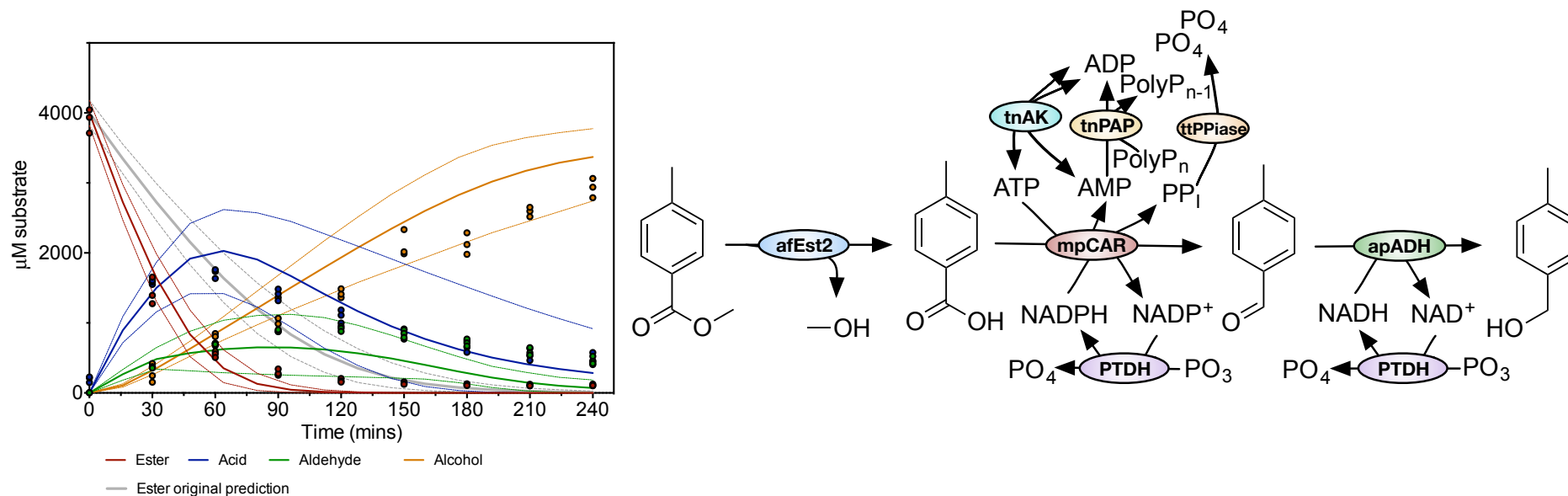
For most industrial processes, the cost of inputs is a critical factor in determining the economic feasibility of an approach. We therefore aimed to demonstrate that this multistep cascade could be optimised using a model to deliver enhanced performance, at reduced input cost. A cost of \$100 per kg of enzyme was chosen, assuming crude lysate is used and enzyme yield has been optimised (Tufvesson *et al.*, 2011). The amount of each enzyme in the pathway was then optimised. The optimisation was required to achieve the lowest price whilst meeting a target of 90% conversion to the alcohol. Substrate concentrations were set at 4,000  $\mu\text{M}$  methyl *p*-toluate, 1,250  $\mu\text{M}$  ATP, 500  $\mu\text{M}$  NADPH, 500  $\mu\text{M}$  NADH, 20,000  $\mu\text{M}$  phosphite and 7,500  $\mu\text{M}$  polyphosphate. A custom built genetic algorithm was used to carry out the optimisation (Figure 5-8). Enzyme concentrations of 9.65  $\mu\text{M}$  Esterase, 0.73  $\mu\text{M}$  CAR, 24.75  $\mu\text{M}$  ADH, 0.07  $\mu\text{M}$  PPIase, 0.43  $\mu\text{M}$  PTDH 0.15  $\mu\text{M}$  AK and 0.57  $\mu\text{M}$  PAP were identified as the lowest cost pathway capable of achieving 90 % yield in four hours.

To test the optimised pathway, reactions were set up at these concentrations. Again the esterase reaction proceeded at approximately twice the expected rate (Figure 5-9). By taking this enhanced rate of esterase activity into account, the model fitted the rest of the data well. Alcohol production was slightly slower than expected reaching a final yield of only 3,000  $\mu\text{M}$  as opposed to the 3,500  $\mu\text{M}$  predicted by the model. However, the optimised pathway performed significantly better in both productivity and cost compared to the non-optimised pathway.



**Figure 5-8 - Flow diagram for genetic algorithm used to optimise the reaction**

A custom built genetic algorithm was used in the optimisation of a batch reaction, minimising total enzyme cost whilst achieving a target yield of 90 % or above. The flow diagram describes the steps the genetic algorithm carries out to reach this goal.



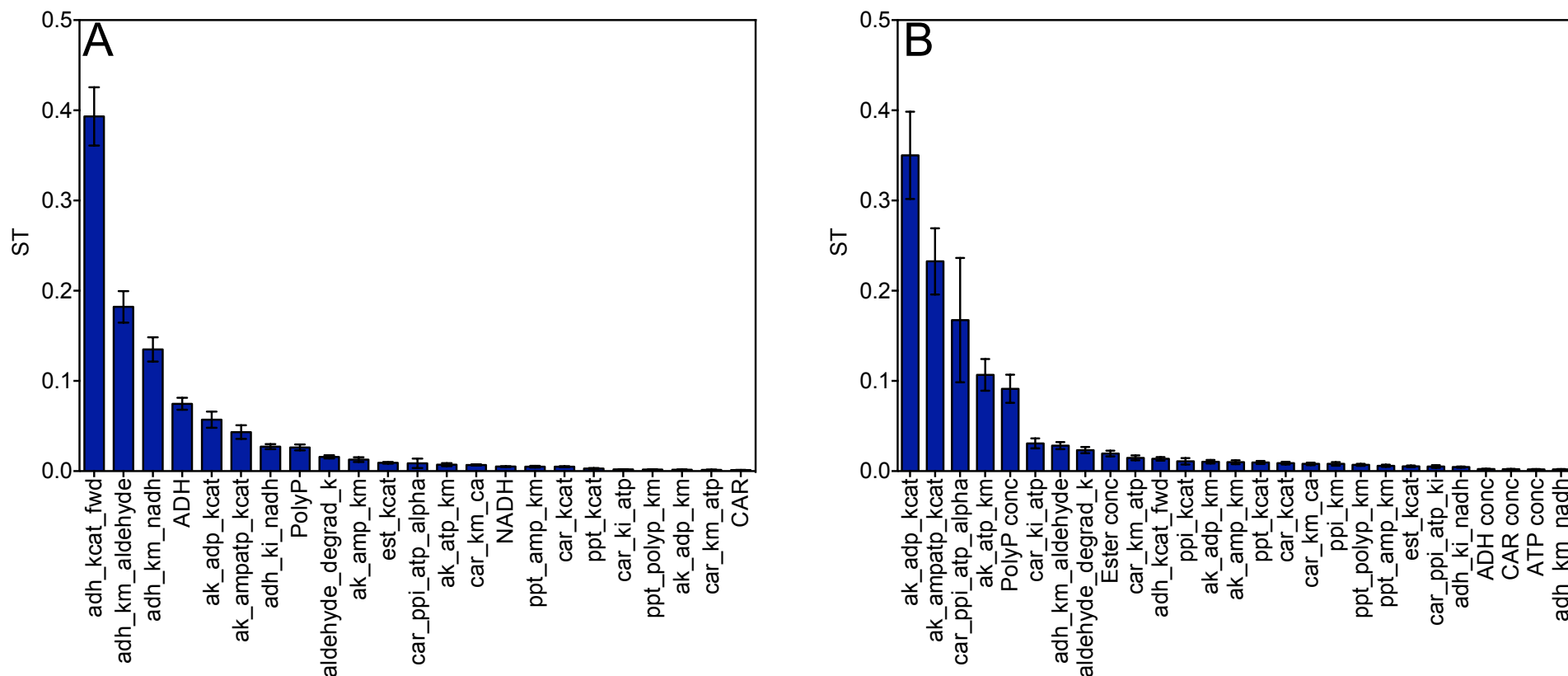
**Figure 5-9 - Validations of the optimised complete reaction**

A reaction using enzyme concentrations predicted by a genetic algorithm to give the lowest cost reaction, whilst hitting a target of 90 % alcohol yield, was performed. Ester, acid, aldehyde and alcohol concentrations measured every 30 minutes are shown as red circles, blue squares, green upwards facing triangles and orange downward facing triangles respectively. The model prediction is shown as the solid line in the same colours. The original model prediction for the ester is shown in grey. Dashed lines represent the 5<sup>th</sup> and 95<sup>th</sup> percentile of the uncertainty analysis. Reactions were initiated with 9.65  $\mu\text{M}$  afEst2, 0.73  $\mu\text{M}$  mpCAR, 0.07  $\mu\text{M}$  ttPPiase, 0.43  $\mu\text{M}$  PTDH, 0.57  $\mu\text{M}$  tnPAP, 0.15  $\mu\text{M}$  tnAK, 4,000  $\mu\text{M}$  methyl *p*-toluate, 500  $\mu\text{M}$  NADPH, 500  $\mu\text{M}$  NADH, 1,250  $\mu\text{M}$  ATP, 20,000 phosphite, 6,000  $\mu\text{M}$  polyphosphate, 20,000  $\text{MgCl}_2$

### **5.3.15 Sensitivity analysis comparison of the optimised vs non-optimised pathways**

As well as delivering optimal performance at lowest cost, it is important that industrial processes are robust, and not easily affected by issues with a single reagent.

Therefore, a Sobol sensitivity analysis was carried out on the model for the optimised reaction to determine the main sources of the large uncertainty in the final alcohol concentration (Nossent, Elsen and Bauwens, 2011) (Figure 5-10). This revealed that the parameters for  $tnAK$  had by far the largest impact on the uncertainty in this optimised reaction. Other parameters with a substantial contribution included CAR inhibition by  $PP_i$ , the rate of aldehyde degradation and polyphosphate concentration. These parameters were very different from those observed in the non-optimised reaction (Figure 5-10). This indicated that, whilst the optimisation had been highly successful in delivering greater productivity at lower cost, this came at the cost of increased sensitivity to some parameters related to individual enzymes.



**Figure 5-10 - Sensitivity analysis of the modelled un-optimised and optimised batch reaction**

The total sensitivity indices (ST) are shown which take into account 1<sup>st</sup> order and all other interactions. Sensitivity is in reference to the uncertainty in the final *p*-tolyl alcohol concentration. Error bars show the 95 % confidence intervals. The sum of all sensitivity indices should equal 1. Parameters with a ST of less than 0.001 are not shown. A: Sensitivity analysis of the un-optimised complete reaction, shown in figure 6C. B: Sensitivity analysis of the optimised complete reaction, shown in Figure 5-9.



## 5.4 Discussion

### 5.4.1 Multi-step *in vitro* biocatalysis

Isolated enzymes are one of the two main systems for industrial biocatalysis, offering several advantages. In particular, they allow reactions to be easily controlled, and so optimized quickly and easily (Hold and Panke, 2009; Bujara *et al.*, 2010; Xue and Woodley, 2012; du Preez *et al.*, 2015). We have demonstrated the use of CARs for multi-step *in vitro* biocatalysis, making use of mechanistic modelling to understand the dynamics of multiple enzymes working in concert. The modular nature of these models allows a systems biology approach to be taken for the creation of new cascades, with the effects of enzyme module addition or removal predictable *in silico* (du Preez *et al.*, 2015).

A critical paradigm for synthetic biology is that “What I cannot create, I do not understand” (Richard Feynman; quote was incorporated into the first chemically synthesized genome (Gibson *et al.*, 2010)). For the synthetic biologist, the building, testing and modelling of biological parts is essential to our understanding of biology (Benner, Yang and Chen, 2011). The creation of a mathematical model for this seven enzyme *in vitro* cascade reaction facilitates its understanding. It acts as a summary for all the available process knowledge for the reaction and allows it to be exploited for the evaluation of different process options, identification of bottlenecks and optimization of the reaction (Woodley, 2013; Ringborg and Woodley, 2016).

### 5.4.2 Modelling drives hypothesis generation

A most significant benefit of mechanistic modelling over a data driven approach is the ability of this form of modelling to generate new hypotheses and understanding of the system (Vasić-Rački, Findrik and Vrsalović Presečki, 2011). In this pathway, modelling was critical in realizing that the buildup of PP<sub>i</sub> slows the CAR reaction significantly. When our model for CAR activity differed from observation, we were able to design new experiments to understand why. Product inhibition of CARs was discovered and characterized (Finnigan *et al.*, 2016), allowing the problem to be solved with the addition of a PPIase enzyme (Teplyakov *et al.*, 1994). This may not have been apparent had data driven modelling been used. We were then able to

test the addition of the PPIase enzyme *in silico* and model that this restored the desired behavior solved the issue before committing to the time consuming process of cloning, expressing and characterizing the enzyme for inclusion in the cascade.

Testing our observations against model predictions revealed some other interesting phenomena. The inclusion of PTDH allowed the CAR reaction to proceed much more effectively than the model predicted at low NADPH concentrations. More detailed modelling would be required to fully understand this; however, this observation might be a result of substrate shuttling or channeling between PTDH and the reductase domain of the CAR enzyme (Tullman-Ercek, 2015). It could be expected that the local concentration of NADP<sup>+</sup> around the CAR enzyme might be higher, and that here PTDH would work more efficiently providing a larger local supply of NADPH (Zheng *et al.*, 2011). Possibly PTDH could even localize near the CAR for this reason. Furthermore, the reaction catalyzed by afEst2 appears to proceed more quickly when included in the full cascade. Possibly afEst2 could be activated by a compound such as MgCl<sub>2</sub>, PO<sub>3</sub>, polyphosphate or NAD(P)H (Tomlinson, Mutus and McLennan, 1981), not included during characterization of afEst2 or the trial batch reaction (Figure 5-4). afEst2 is an unusual esterase in that it has a Coenzyme A molecule tightly bound near its active site, which could be playing a regulatory role, possibly modulated by one of these additional compounds (Sayer *et al.*, 2016). Alternatively substrate shuttling between afEst2 and mpCAR could be occurring, resulting in a faster afEst2 catalyzed reaction as the product removed more quickly (Tullman-Ercek, 2015).

#### **5.4.3 The use of CARs *in vitro***

CARs have previously been used in whole-cell biocatalysis where ATP and NADPH can be regenerated by the host metabolism (Akhtar, Turner and Jones, 2013). PP<sub>i</sub> inhibition has likely not been identified as a problem for CAR enzymes previously, due to the activity of native PPIase enzymes (Kankare *et al.*, 1994). However while the use of whole-cell biocatalysis for CAR reactions has its advantages, the optimization of whole-cell systems is challenging, and side reactions can be problematic. For example, native ADH enzymes may cause reduction all the way to the alcohol product even when this isn't desired (France *et al.*, 2016). The use of a CAR as an isolated enzyme allows a more efficient process to be designed, and the

inclusion of cofactor regeneration provides an economically competitive alternative to whole-cell biocatalysis (Zhao and Van Der Donk, 2003). Importantly, modelling the use of a CAR *in vitro* has allowed a more thorough understanding of its reaction kinetics, useful not only *in vitro*, but also for informing *in vivo* design (Rios-Solis *et al.*, 2015).

#### 5.4.4 Modelling NADH and NADPH regeneration

PTDH accepts two substrates,  $\text{NAD}^+$  and  $\text{NADP}^+$ , which are competing against each other for the enzyme's active site (Woodyer, Van der Donk and Zhao, 2003).

Increasing the concentration of one will increase the apparent  $K_M$  for the other, as they are competitive inhibitors of each other (the reaction was modelled as such). In our optimized reaction both  $\text{NAD}^+$  and  $\text{NADP}^+$  were set to 500  $\mu\text{M}$  and only PTDH concentration optimized. In this situation  $\text{NAD}^+$  is regenerated more effectively as PTDH has a lower  $K_M$  and higher  $k_{cat}$  for this substrate. Possibly, the PTDH reaction could be optimized further by allowing the ratio of  $\text{NAD}^+ : \text{NADP}^+$  to be altered in the optimization, such that rates of NADH and NADPH regeneration are balanced to the requirements of the reaction. The reaction might be further improved by considering the evidence for substrate shuttling of cofactors between enzymes, possibly allowing expensive cofactor concentrations to be reduced significantly (Kragl *et al.*, 1996). This would require substantial additional modelling, but would offer considerable advantages for *in vitro* biocatalysis.

#### 5.4.5 The ATP regeneration system is challenging to model

AK catalyzes an equilibrium between  $\text{MgADP} + \text{ADP} \leftrightarrow \text{MgATP} + \text{AMP}$ . The free magnesium concentration is critical in regulating the position of this equilibrium (Blair, 1970). While adenylate kinase has been shown to follow a random Bi Bi mechanism (Sheng, Li and Pan, 1999), in practice this is difficult to implement into a mathematical model as free magnesium concentration would also need to be modelled. To further complicate this modelling, polyphosphate strongly chelates magnesium ions depending on the concentration and length of the polyphosphate chains, both of which can only be estimated (Strauss and Siegel, 1963). Because of these challenges we were forced to only approximate the AK reaction using a bi substrate equation (Table 5-1), with high levels of uncertainty associated with each

parameter. Previous work has suggested where the equilibrium constant for this reaction might lie, at a range of free magnesium concentrations (Blair, 1970). These values are easily captured within the parameter bounds set for the uncertainty analysis of our approximation, indicating that our modelling is biochemically reasonable. However, when AK concentrations were optimized, this increased the overall uncertainty in the modelling significantly. It may be that the optimization strategy should be altered to use more than the minimum levels of enzymes with high uncertainties, to increase the robustness of the overall model.

The reaction catalyzed by PAP is also difficult to model. Initially it was hoped that polyphosphate concentration could be set at a high level such that its effects on reaction rate could be ignored. However, polyphosphate strongly chelates  $\text{Mg}^{2+}$ , and  $\text{MgCl}_2$  cannot be added at a high concentration due to the inhibitory effects this would cause on other enzymes (Strauss and Siegel, 1963). This limited feasible polyphosphate concentrations to approximately 10 mM or less, with the inclusion of 20 mM  $\text{MgCl}_2$ . Previous characterization of PAP enzymes is limited, and the reaction mechanism unknown (Bonting, Kortstee and Zehnder, 1991; Resnick and Zehnder, 2000; Kameda *et al.*, 2001; Itoh and Shiba, 2004; Shiba *et al.*, 2005). PAPs are thought to be processive, as tri- or di- phosphates are not accepted by the enzyme, yet these are never detected following a reaction (Kameda *et al.*, 2001). Both polyphosphate chain length and concentration affect reaction kinetics, making this challenging to model accurately (Shiba *et al.*, 2005). Because of this we used a bi substrate equation to approximate the PAP reaction. We acknowledge that this reduces the robustness of the overall model: again, this suggests that an optimized pathway might use more than the minimum concentration of this enzyme to mitigate this.

#### **5.4.6 The optimised reaction**

Optimization of the batch reaction used to validate the model demonstrates its potential for exploring new process options. The optimization of the reaction resulted in just enough AfEst2 to hydrolyze all the methyl *p*-toluate in the time available, and just enough ApADH to produce 90 % *p*-tolyl alcohol within 4 hours (Figure 5-9). In contrast, *p*-toluic acid and *p*-tolualdehyde concentrations were maintained at a low but steady concentration, allowing maximum productivity by all enzymes for the

entire reaction. Enzyme concentrations were minimized and productivity maximized successfully.

#### 5.4.7 Comparing the optimised and un-optimised reactions.

In comparing the uncertainty in the un-optimized (Figure 5-7C) and optimized (Figure 5-9) complete reactions, it is clear that uncertainty in the optimized reaction has increased significantly. Minimizing the concentration of all enzymes so that they are each close to being rate limiting has likely increased the uncertainty in substrate concentrations, as changes in a greater number of parameters can impact on the rates.

The initial reaction still showed reasonably large uncertainty in *p*-toluic acid and *p*-tolualdehyde concentrations. However, these do not translate to a high degree of uncertainty in the *p*-tolyl alcohol concentration. When *p*-tolualdehyde concentration is high enough (well above its  $K_M$  for ApADH), the main cause of uncertainty in *p*-tolyl alcohol concentration is the rate of the ApADH reaction (primarily `adh_kcat_fwd`; Figure 5-10A). However, when *p*-tolualdehyde is maintained at a lower concentration (as in the optimized reaction), *p*-tolualdehyde concentration has a greater impact on pathway efficiency. In this case many of the parameters causing uncertainty in *p*-tolualdehyde concentration are also responsible for the larger degree of uncertainty in *p*-tolyl alcohol concentration, as shown in the sensitivity analysis (Figure 5-10B). If *p*-tolualdehyde concentration were to be maintained well above its  $K_M$  for apADH, *p*-tolyl alcohol production should only be sensitive to apADH parameters, regardless of uncertainty in the other steps. However as *p*-tolualdehyde is shown not to be stable in the reaction (Figure 5-5), this compound is best kept at a low concentration.

Future optimization could include an additional objective for the minimization of uncertainty in the final *p*-tolyl alcohol concentration. This would allow cost to be minimized whilst also allowing a robust process to be designed. However, the inclusion of two competing objective functions would make the optimization process more complex, likely requiring the generation and manual evaluation of a set of Pareto optimal solutions (Deb *et al.*, 2000).

#### 5.4.8 Evaluating other potential reaction configurations

For the use of isolated enzymes to be economical, the reuse of the enzymes for multiple reactions should be considered. The immobilization of enzymes is a powerful approach facilitating this (Sheldon and Van Pelt, 2013). The use of membranes with low molecular weight cut-offs can allow an enzyme cocktail to be maintained while substrates and products can be easily added and removed. An interesting example might be the use of an enzyme “teabag” (Wachtmeister *et al.*, 2014). Diffusion across a membrane could be included in the model to help evaluate this option.

Flow reactors are another popular option, in which immobilized enzymes (such as Cross-Linked Enzyme Aggregates - CLEAs (Sheldon, 2011)) are packed into reaction channels through which substrates and products flow (Hickey *et al.*, 2007). For the construction of multistep reactions, using flow reactor modelling will be invaluable in informing the set-up of the reactor, allowing high productivity and conversion to be achieved. A two phase system is another alternative (Woodley and Lilly, 1990). Hydrophobic solvents can be used as a substrate and product sink for hydrophobic substrates, allowing high substrate concentrations to be used (Domínguez de María and Hollmann, 2015). Diffusion between the phases could be modelled helping inform solvent selection and reaction set up.

Isolated enzymes are often used as a crude lysate (Tufvesson *et al.*, 2011). An interesting approach for the construction of this multi enzyme reaction might be the expression of all seven enzymes in one host organism. Expression levels could be balanced using synthetic biology techniques, for example using characterized ribosome binding sites to modulate expression of each enzyme in a single operon. The level of expression required could be determined using the model, to design an organism with the correct expression levels of each enzyme. If a thermostable CAR enzyme were used, purification by differential thermal denaturation would also be possible (Ninh *et al.*, 2015).

#### 5.4.9 Operational stability

Data for enzyme degradation under operating conditions could be included in the model to allow evaluation of the reaction for extended operation. This is an

important consideration in considering the overall cost and productivity of a reaction and could be included in the model (Vasić-Rački, Findrik and Vrsalović Presečki, 2011). Enzymes from thermophiles have been used as they are an excellent starting point in achieving high operational stability; although mpCAR is the weakest enzyme in this respect, it still has a long half-life at 30 °C (Finnigan *et al.*, 2016). The use of enzyme immobilisation or the identification of a more thermostable CAR could improve on the operational stability of the whole reaction. The identification of the operational window for the reaction (Figure 5-3), is also an important step in designing an optimal reaction with good operational stability (Xue and Woodley, 2012).

#### **5.4.10 Conclusions**

Multistep reactions are highly attractive to industry, with enzymes ideally suited for these processes given their ability to operate under similar reaction conditions, and offering specificity and selectivity in their reactions. Here, we show that mechanistic modelling of a three-step reaction pathway with seven enzymes is not only possible, but faithfully predicts experimental results. Furthermore, we demonstrate that the use of a mechanistic model allows bottlenecks and deviations from predicted behavior to be identified, providing opportunities to rapidly overcome blockages. Our model allowed optimization of the reaction pathway to provide higher throughput at lower cost, highlighting the potential for application to industry.

By using good modelling practice and considering the uncertainty of each step in the pathway (Sin, Gernaey and Lantz, 2009), we show which steps and parameters are of greatest importance to the reaction, guiding improvements to reaction robustness. Uncertainty analysis also facilitated the use of approximations where reactions were too prohibitively complex to model entirely faithfully, with the error associated with these approximations captured in the uncertainty and sensitivity analysis.

Our model reaction is highly adaptable to a wide range of reactions, as the central CAR model is ideal for linking together pharmaceutically important chemical pathways. An enzyme toolbox featuring well characterized enzyme parts and covering a network of reactions will drive multi-step biocatalysis process development forward. The construction of models for enzymes in the toolbox allows

novel multi-step reactions to be quickly tested *in silico*, with the construction of a model early in the development process an invaluable asset.

A parallel with the design, model, build and test cycle of synthetic biology may be drawn for the process of model construction, and the use of validated parts for the construction of future reactions both *in vitro* and *in vivo*. The process of building and testing a model can drive hypothesis generation, and can be an invaluable exercise in developing a new process.

## 5.5 Methods

### 5.5.1 Materials and plasmids

All chemicals were purchased from Sigma-Aldrich (Gillingham, UK), and were of the highest purity available. Plasmid details are available as supplementary file 2.

### 5.5.2 Enzyme preparation

Genes for the expression of mpCAR, PTDH, tnAK, tnPAP, and ttPPIase were cloned into the *Nco*I and *Hind*III sites of pNIC28-Bsa4(Hansch *et al.*, 1977). The native ApADH gene sequence was cloned previously into pET-30 Xa/LIC (Hickey, 2008). The gene for mpCAR was obtained by PCR from genomic DNA; all others were codon optimized for *E. coli* and gene synthesized by IDT. All contained a N-terminal 6x histidine tag (Hansch *et al.*, 1977). Vectors were transformed into BL21 (DE3) *E. coli* for expression. mpCAR was co-transformed with a pCDF-Duet1 vector containing a phosphopantetheine transferase from *Bacillus subtilis*. A pET24-d plasmid containing afEst2 transformed into the *E. coli* strain BL21-CodonPlus (DE3)-RIPL (Agilent) has been described previously (Sayer *et al.*, 2016). Expression was carried out in LB media with the addition of 50 µg/µl of appropriate antibiotics.

Cells were grown to approximately 0.6 OD<sub>600nm</sub> at 37 °C with shaking at 225 rpm, at which point IPTG was added to a concentration of 100 µM. Temperature was dropped to 20 °C for protein expression overnight, except in the case of afEst2 where temperature was dropped to 30 °C. Cells were harvested by centrifuging and re-suspended in 25 mM Tris-HCl pH 8.0, 0.5 M NaCl<sub>2</sub>.



Cell lysate was prepared by sonication on ice and clarified by centrifugation at 20,000g at 4 °C. Enzymes were purified from the cell lysate using a 1 ml His-Trap FF crude column (GE Healthcare) using an elution gradient from 10 to 250 mM imidazole in 25 mM Tris-HCl pH 8.0, 0.5 M NaCl. The purified sample was then applied to a Superdex 200 HiLoad 16/60 size exclusion column (GE Healthcare) and eluted in 25 mM HEPES pH 7.5, 0.1 M NaCl at 1.0 ml/min. Eluted fractions were analyzed by SDS-PAGE before being pooled and concentrated to between 1 and 8 mg/ml. To calculate protein concentration from OD<sub>280nm</sub>, an extinction coefficient and molecular weight for each enzyme was calculated using the ExPaSy ProtParam tool. Single use aliquots of protein were stored at -80 °C.

### **5.5.3 Enzyme assays**

Unless otherwise stated, all reaction were performed in triplicate in a 96-well microtitre plate using a Tecan M200 Infinite plate reader. All assays were carried out in a standard reaction buffer consisting of 100 mM Tris-HCl at pH 7.5, titrated to the correct pH whilst at 30 °C.

### **5.5.4 Buffers for pH vs activity assays**

Buffers to measure activity across a range of pH values were prepared at 0.2 pH unit intervals at assay temperature. Buffers used were 50 mM citrate between pH 4.0 and 5.8, 50 mM MES between pH 5.8 and 6.6, 50 mM PIPES between pH 6.4 and 7.4, 50 mM MOPS between pH 6.8 and 7.8, 50 mM HEPES between pH 7.0 and 8.0, 50 mM Tris between pH 7.8 and 9.0, and boric acid between pH 8.8 and 10. HCl or NaOH was used to titrate the buffers to the correct pH as appropriate. In some cases data were adapted from previous work as indicated in the relevant sections.

### **5.5.5 Thermostability assays**

Thermostability was measured by heating enzyme samples for half an hour using the temperature gradient of a Bio-Rad thermocycler between 30 and 90 °C (or less where appropriate) before cooling on ice. Activity was then measured using specific enzyme assays relative to a control sample kept on ice. In some cases data were adapted from previous work as indicated in the relevant sections.

### 5.5.6 CAR assays

To determine the two substrate kinetics of mpCAR with ATP and *p*-toluic acid, reactions were set up containing 11  $\mu\text{g}$  / ml mpCAR enzyme, 0.25 mM NADPH, 10 mM  $\text{MgCl}_2$ , with a range of ATP and *p*-toluic acid concentrations. The oxidation of NADPH was used to monitor reactions by measuring absorbance of NADPH at  $\text{OD}_{340\text{ nm}}$ ; production of aldehyde, measured by HPLC, correlates tightly with the NADPH oxidation (Finnigan *et al.*, 2016). Reactions were carried out for 10 minutes after a 5 minute pre-incubation at 30 °C. The appropriate equation was determined by fitting the initial rates of reaction at eight ATP concentrations around the expected  $K_M$  including a blank, each at five *p*-toluic acid concentrations around its expected  $K_M$ . Data were fit by least squares non-linear regression using GraphPad Prism 5.0, and possible reaction mechanisms compared for goodness of fit. A  $K_M$  for NADPH was determined by setting up reactions containing 10  $\mu\text{g}$  / ml mpCAR enzyme, 10 mM  $\text{MgCl}_2$ , 1 mM ATP, 10 mM *p*-toluic acid and a range of eight NADPH concentrations around its expected  $K_M$ , after confirming NADPH concentration at  $\text{OD}_{340\text{ nm}}$ . Other characterization was carried out previously (Finnigan *et al.*, 2016).

### 5.5.7 ADH assays

To determine kinetic parameters for apADH, assays were carried out in sealed PCR tubes using a Bio-Rad thermocycler. After 15 to 30 minutes reactions were quenched by transfer into a 96-well microtitre plate containing 5 mM NaOH. Increase or decrease in NADH concentration was measured at  $\text{OD}_{340\text{ nm}}$  to determine activity, relative to a blank reaction. Reactions to measure activity in the reductive direction were set up containing either 0.5 mM NADH or 10 mM *p*-tolualdehyde. Conversely, reactions to measure activity in the oxidative direction were set up containing either 2.5 mM  $\text{NAD}^+$  or 100 mM *p*-tolualdehyde. In each case a range of concentrations were used for the other respective substrate around its expected  $K_M$ , including a blank. *p*-tolylaldehyde kinetics were measured at 30 °C in standard reaction buffer using 166.7  $\mu\text{g}$  / ml apADH.  $\text{NAD}^+$ , NADH and *p*-tolyl alcohol kinetics were carried out at 70 °C in 100 mM Tris-HCl titrated to pH 7.5 at 70 °C, using 40  $\mu\text{g}$  / ml apADH. Only  $K_M$  constants were taken from data at 70 °C. Data were fit by least squares non-linear regression using GraphPad Prism 5.0. Data for the effects

of pH and temperature were adapted from previous work (Hirakawa, Kamiya and Kawarabayashi, 2004).

### 5.5.8 PTDH assays

To determine kinetic parameters for PTDH, assays were set up containing 20 mM  $\text{Na}_2\text{HPO}_3$ , 2.8  $\mu\text{g}$  / ml PTDH, and a range of  $\text{NAD}^+$  or  $\text{NADP}^+$  concentrations around the expected  $K_M$ . The production of NADH or NADPH was monitored at  $\text{OD}_{340\text{nm}}$  and data were fit by least squares non-linear regression using GraphPad Prism 5.0. The thermostability of a sample containing 18  $\mu\text{g}$  / ml PTDH was measured as described previously. Activity at various pH values was determined using a range of buffers described previously, in place of the standard reaction buffer.

### 5.5.9 PPIase

PPIase activity was measured using the production of phosphomolybdate to measure phosphate content, as described (Heinonen and Lahti, 1981). Using a standard curve, the rate of phosphate production by ttPPIase was determined using a high (5 mM) concentration of  $\text{PP}_i$  in the presence of 10 mM  $\text{MgCl}_2$ , in standard reaction buffer. Five measurements were taken in triplicate over 20 minutes. A conservative  $K_M$  was estimated from the BRENDA database (Schomburg *et al.*, 2013) and the rate of  $\text{PP}_i$  hydrolysis used to calculate  $k_{\text{cat}}$ .

Activity at pH was determined using the same phosphomolybdate method with reaction buffers covering a range of pH values, detailed previously. The thermostability of a sample containing 0.44  $\mu\text{g}$  / ml ttPPIase was measured as described previously, with the assay above used with readings at time points 0 and 20 minutes taken. 10 mM  $\text{MgCl}_2$  was included in the heated sample.

### 5.5.10 PAP assays

ADP formation was measured using an ADP-Glo kinase assay kit (Promega) in a 384-well solid white microtitre plate (Corning, 3572) following manufacturer's instructions. Assays were carried out in standard reaction buffer containing 20 or 40 mM  $\text{MgCl}_2$ , 12.5 mM polyphosphate, 3.25 mM AMP, 12.5  $\mu\text{g}$  / ml or 125  $\mu\text{g}$  / ml of tnPAP enzyme. Reactions were carried out for 15 minutes before quenching with the ADP-Glo kit. A range of polyphosphate, AMP and  $\text{MgCl}_2$  concentrations around

an expected  $K_M$  were assayed in turn. ADP production was calculated from a standard curve, and the data fit by least squares non-linear regression using GraphPad Prism 5.0. Specific assay conditions are shown in Table 5-2 - Supplementary Figure 16 –Table 5-2 - Supplementary Figure 20. Activity at pH was measured using a range of buffers described previously in place of the standard reaction buffer. Thermostability was determined as described previously, using the ADP-Glo assay to measure residual relative activity.

### **5.5.11 Mathematical modelling**

Mathematical modelling was carried out in Python 3.4 using the SciPy and SALib modules. Outputs were exported to GraphPad Prism 5.0 for plotting. The `integrate.odeint` function in SciPy was used to solve ordinary differential equations.

Uncertainty analysis was carried out making use of the SALib module. Input bounds for each parameter were defined as either the calculated 95 % confidence intervals, or where uncertainty was judged to be high by 25 or 50 % of the parameter value, as shown in Table 5-2. Sampling of the possible inputs within these bounds was carried out by Latin hypercube sampling of 1,000 samples, with the model for each sample run. At each time point the mean, 5<sup>th</sup> and 95<sup>th</sup> percentile of each substrate concentration was plotted to represent the uncertainty of the model.

The method of Sobol was used for sensitivity analysis as part of the SALib module. Second order effects were not calculated, and the total sensitivity indices plotted to show the sensitivity of each parameter. The sample number was set at 1,000.

### **5.5.12 Optimisation using a genetic algorithm**

A custom built, single objective genetic algorithm was used to minimize enzyme cost, on the condition over 90 % alcohol yield was reached. A general outline of the algorithm is shown in Figure 5-8, and the script is provided as a Supplementary File.

### **5.5.13 Model validation reactions**

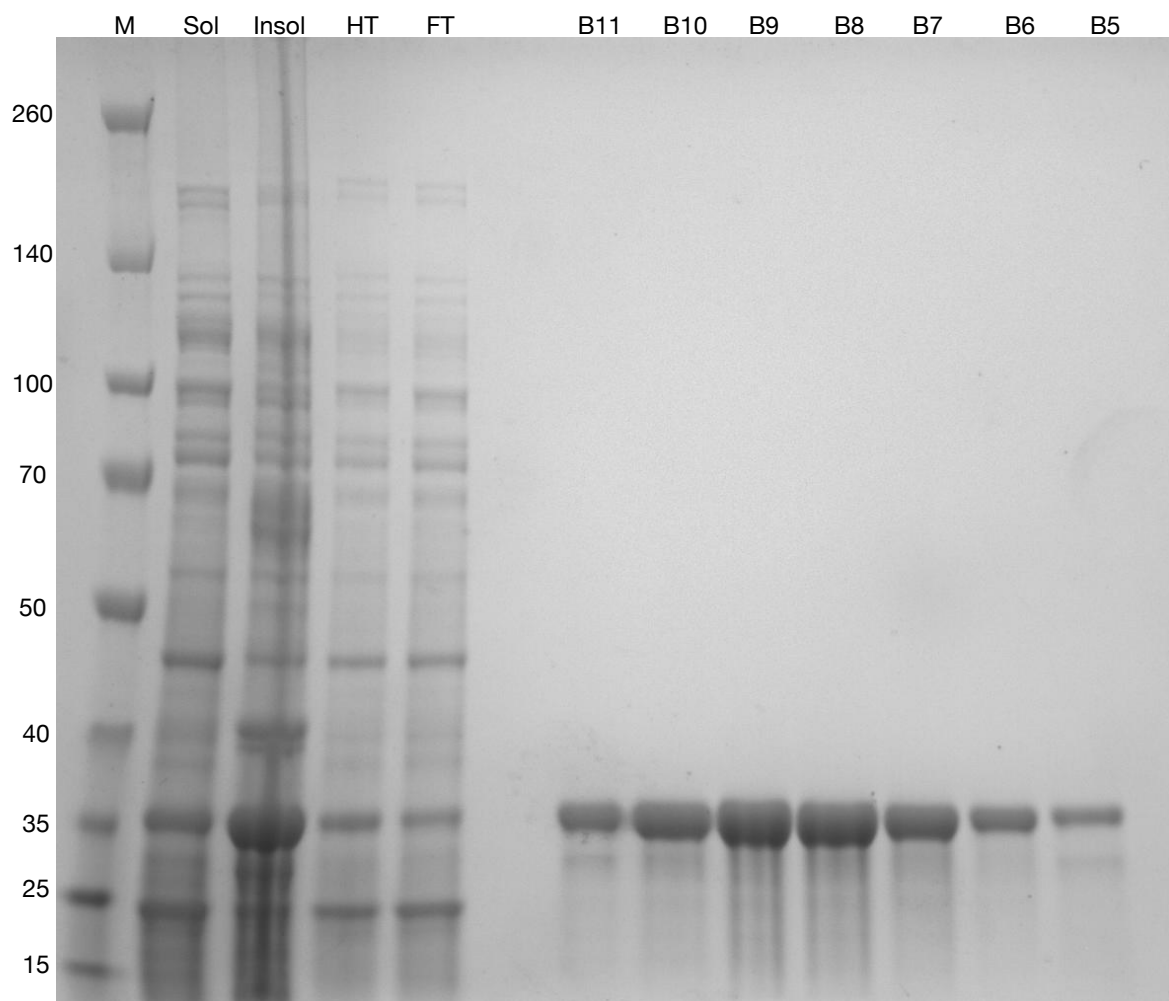
Standard curves of methyl *p*-toluate, *p*-toluic acid, *p*-tolylaldehyde and *p*-tolyl alcohol were prepared and analyzed by HPLC. 500  $\mu$ l reactions were set up in triplicate in 1.5 ml microcentrifuge tubes, and incubated in a thermoshaker at 30 °C with 500 rpm

shaking. 50 µl samples were taken every 30 minutes into 50 µl acetonitrile and centrifuged for 10 minutes. Supernatant was removed and stored at 4 °C until analyzed by HPLC.

#### **5.5.14 HPLC**

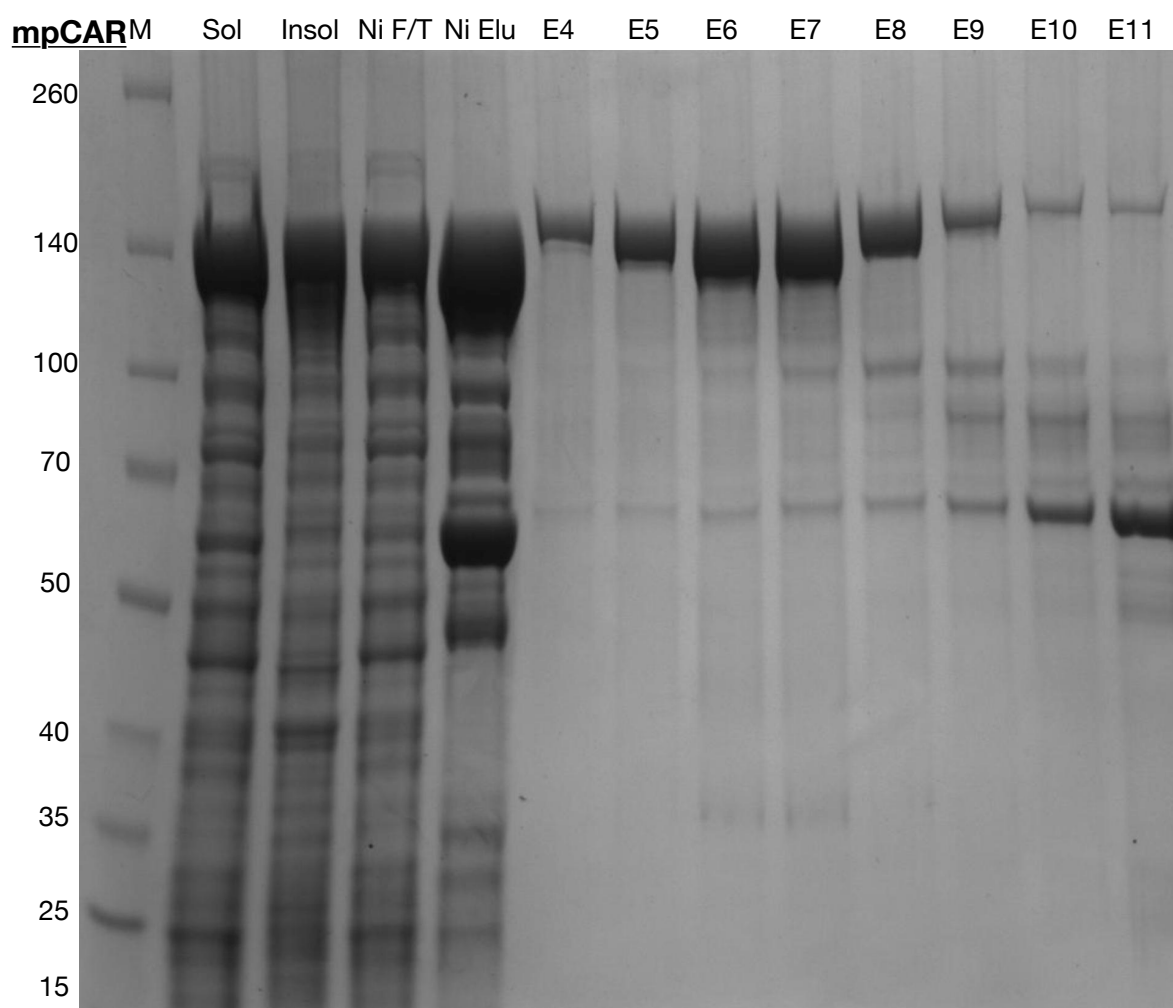
An Eclipse Plus C18 column with a particle size of 3.5 µm, measuring 4.6 x 100 mm, was used. The column was run at 60 °C on the following method using two buffers, buffer A: 95 % H<sub>2</sub>O, 5 % (v/v) acetonitrile, 0.1 % (v/v) trifluoroacetic acid, and buffer B: 5 % H<sub>2</sub>O, 95 (v/v) % acetonitrile, 0.1 % (v/v) trifluoroacetic acid. 3 µl of sample was injected and eluted on a gradient from 0 to 100 % buffer B over 10 minutes. Buffer B was maintained at 100 % for a further 2 minutes before the column was re-equilibrated with buffer A for 2 minutes before the next run.

## 5.6 Supplementary Figures



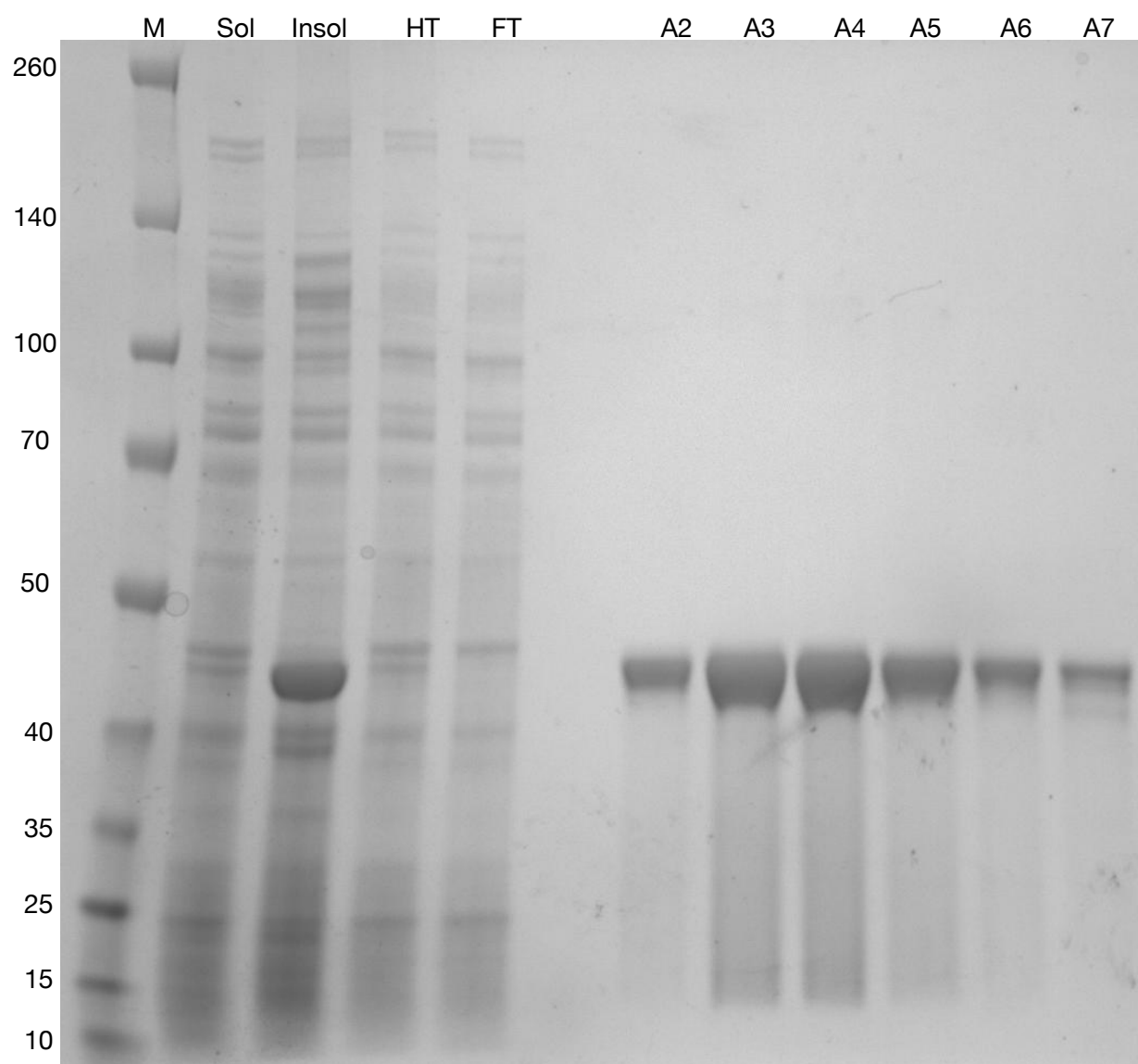
**Figure 5-2 - Supplementary Figure 1 - SDS-PAGE analysis of the AF-Est2 protein purification**

Sol: Soluble fraction of cell lysate. Insol: Insoluble fraction of cell lysate. HT: Soluble lysate heat treated at 70 °C for 30 minutes before removal of precipitated proteins by centrifugation. FT: The flow through after loading the heat treated sample onto the nickel column. B11-B5: Fractions collected from the gel filtration column believed to be the purified AF-Est2 protein with an expected MW of 29kDa.



**Figure 5-2 - Supplementary Figure 2 - SDS-PAGE analysis of the mpCAR protein purification.**

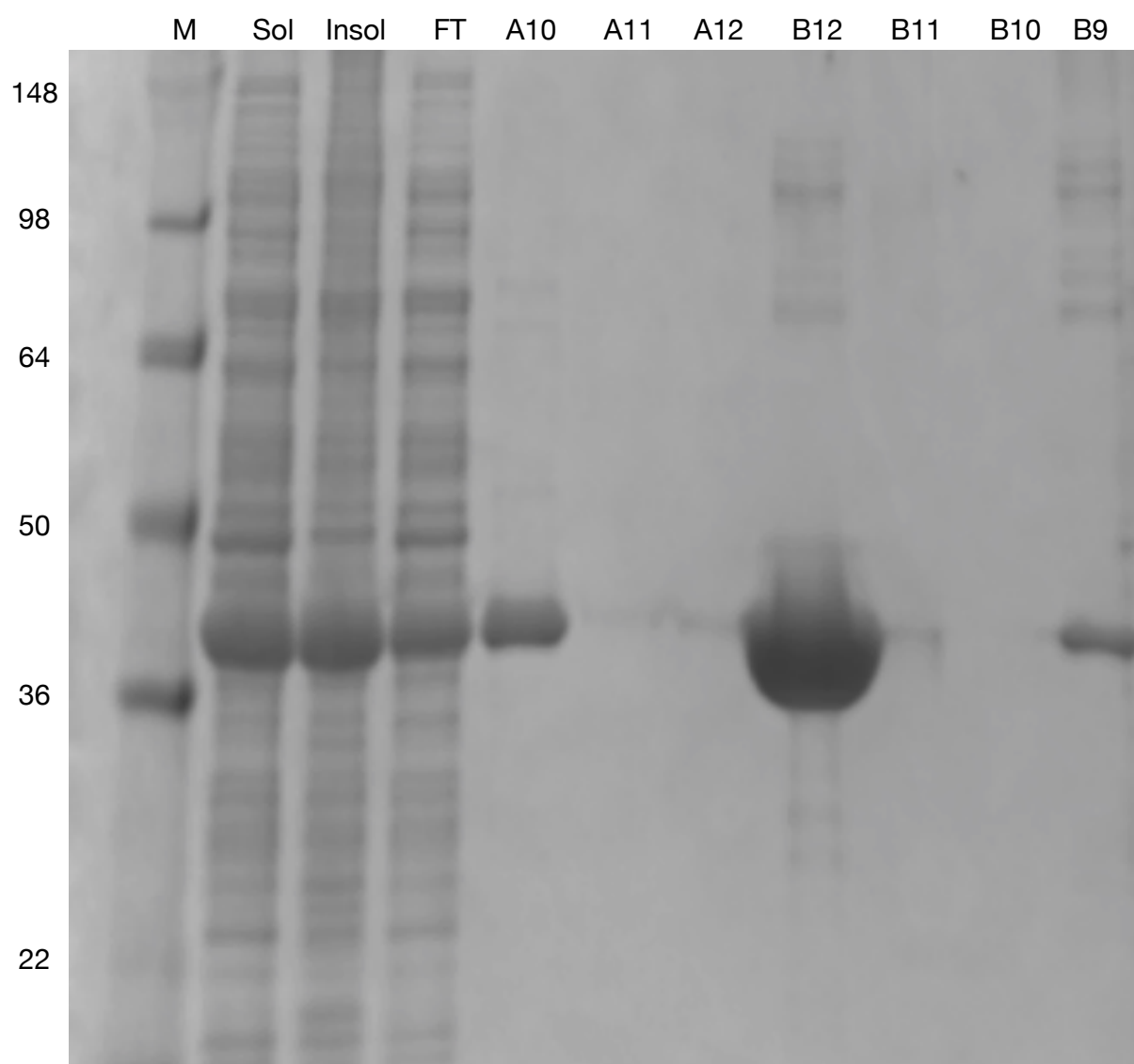
Sol: Soluble fraction of cell lysate. Insol: Insoluble fraction of cell lysate. Ni F/T: The flow through after loading the sample onto the nickel column. Ni Elu: Purified protein after the nickel purification step. E4-E11: Fractions collected from the gel filtration column believed to be the purified mpCAR protein with an expected MW of 128kDa. Fractions E9 – E11 contained another unknown protein at approximately 60kDa. These fractions were not pooled with the remaining fractions.



**Figure 5-2 - Supplementary Figure 3 - SDS-PAGE analysis of the ApADH protein purification.**

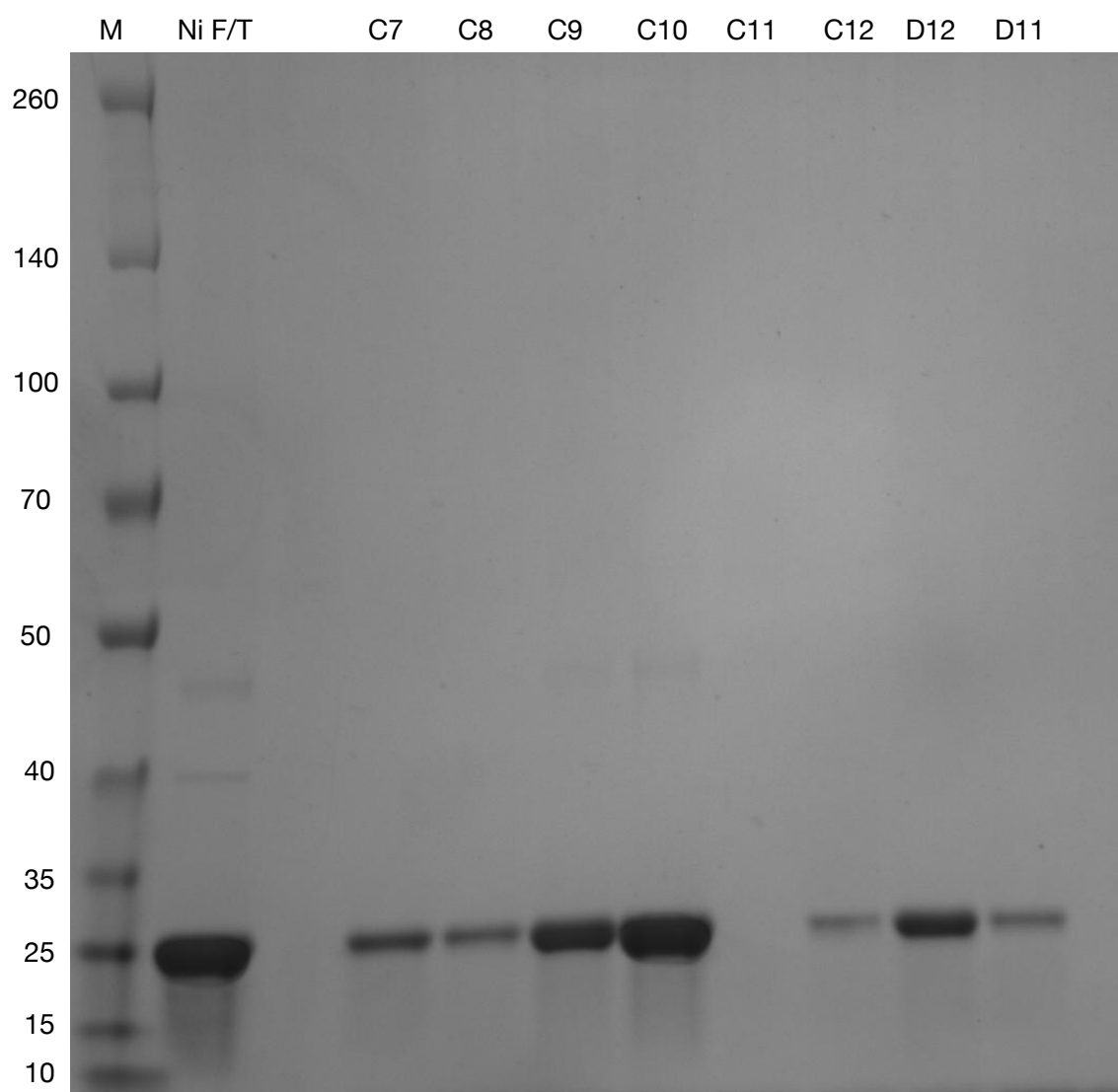
Sol: Soluble fraction of cell lysate. Insol: Insoluble fraction of cell lysate. HT: Soluble lysate heat treated at 70°C for 30 minutes before removal of precipitated proteins by centrifugation. FT: The flow through after loading the heat treated sample onto the nickel column. A2-A7: Fractions collected from the gel filtration column believed to be the purified ApADH protein with an expected MW of 41 kDa.





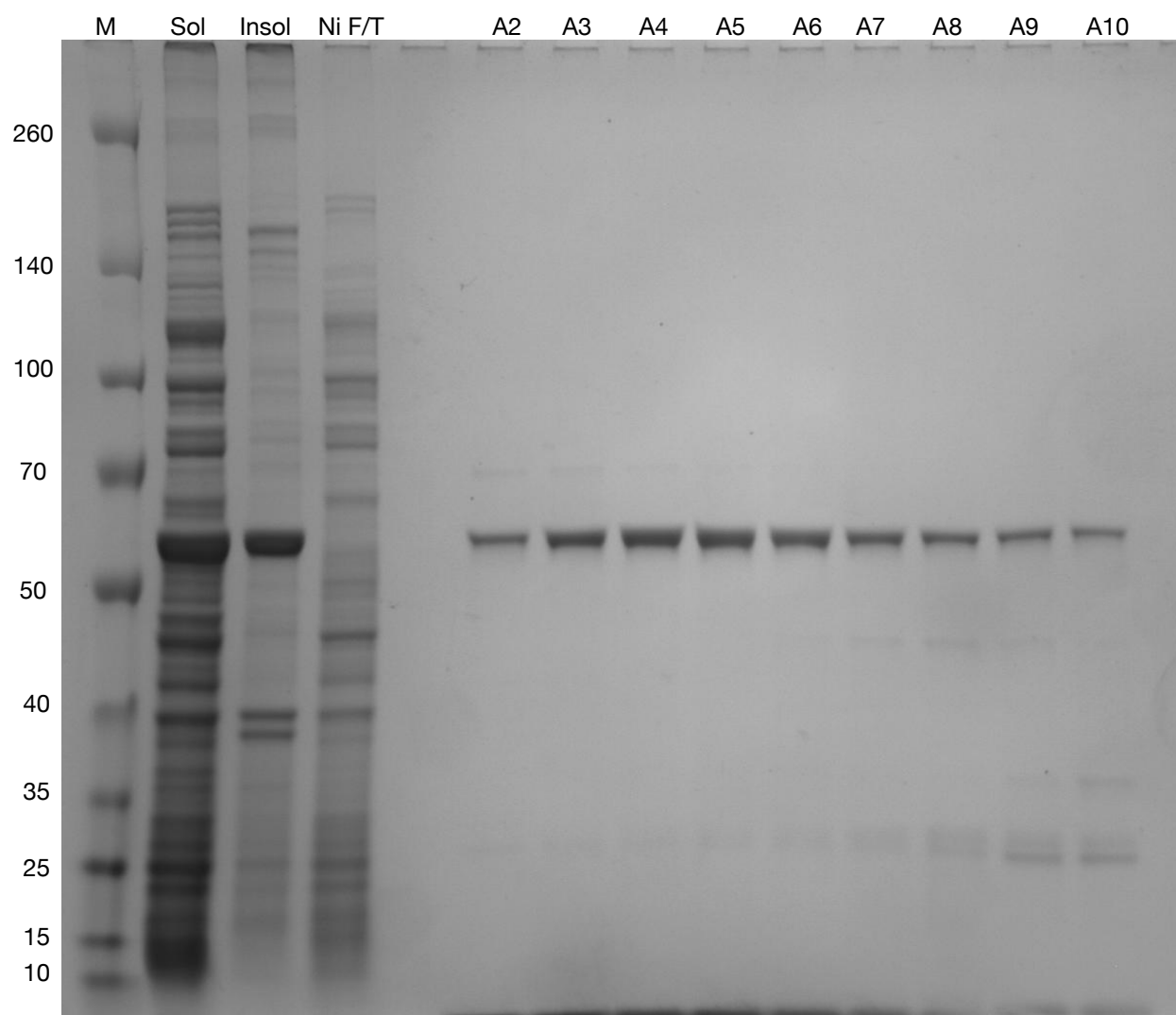
**Figure 5-2 - Supplementary Figure 4 - SDS-PAGE analysis of the PTDH protein purification.**

Sol: Soluble fraction of cell lysate. Insol: Insoluble fraction of cell lysate. FT: The flow through after loading the sample onto the nickel column. A10-B9: Fractions collected from the gel filtration column believed to be the purified PTDH protein with an expected MW of 38.7 kDa. Very little protein was detected by SDS-PAGE in lanes A11, A112, B11 and B10, however there was likely a problem in running in the gel as there was an unavoidable delay between preparing the samples and running them. Samples at the extremes of the collected peak show pure protein so the entire peak was used.



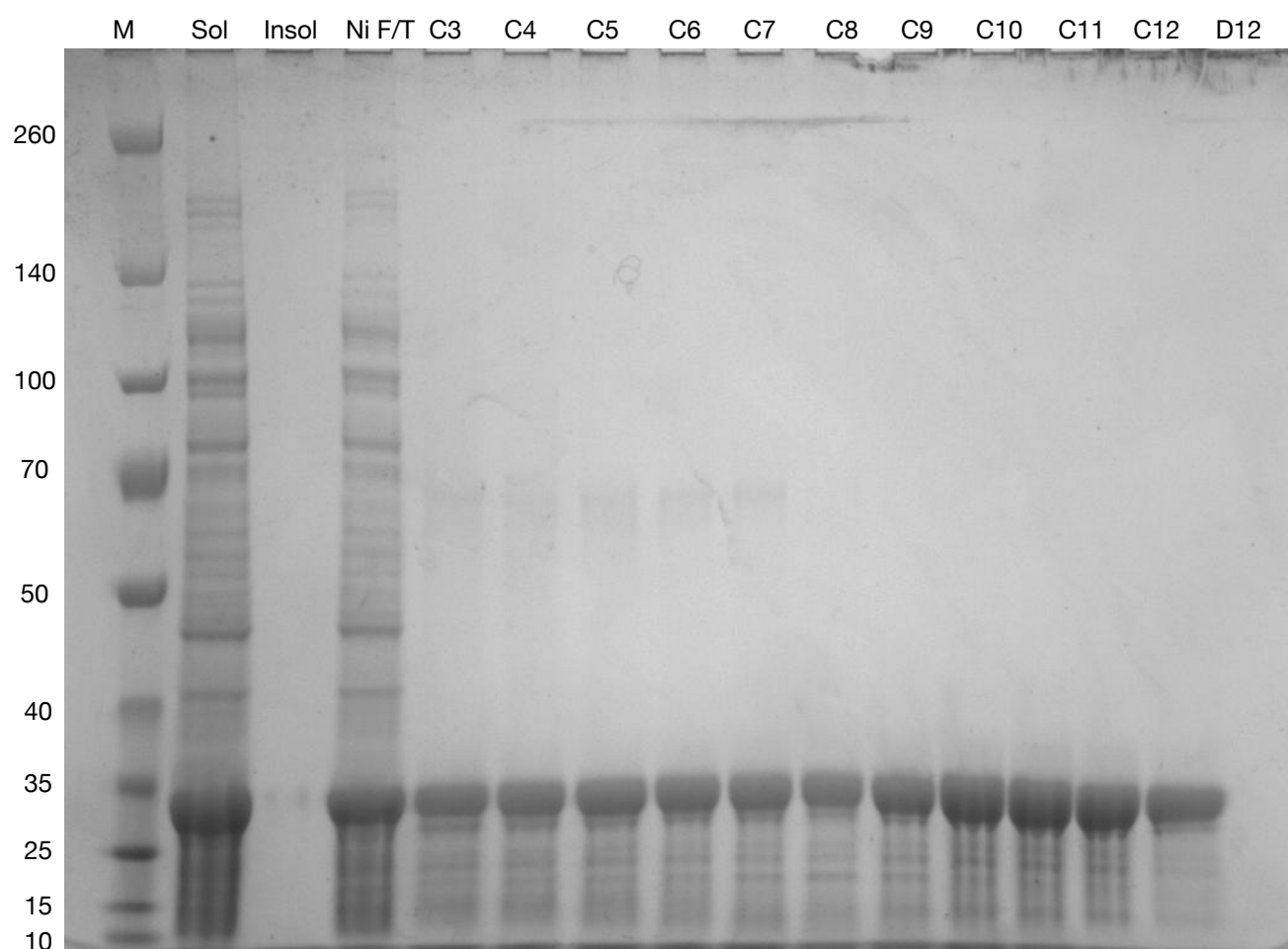
**Figure 5-2 - Supplementary Figure 5 - SDS-PAGE analysis of the ttPpiase protein purification.**

Ni F/T: The flow through after loading the heat treated sample onto the nickel column. C7-D11: Fractions collected from the gel filtration column believed to be the purified ttPpiase protein with an expected MW of 21.8kDa.



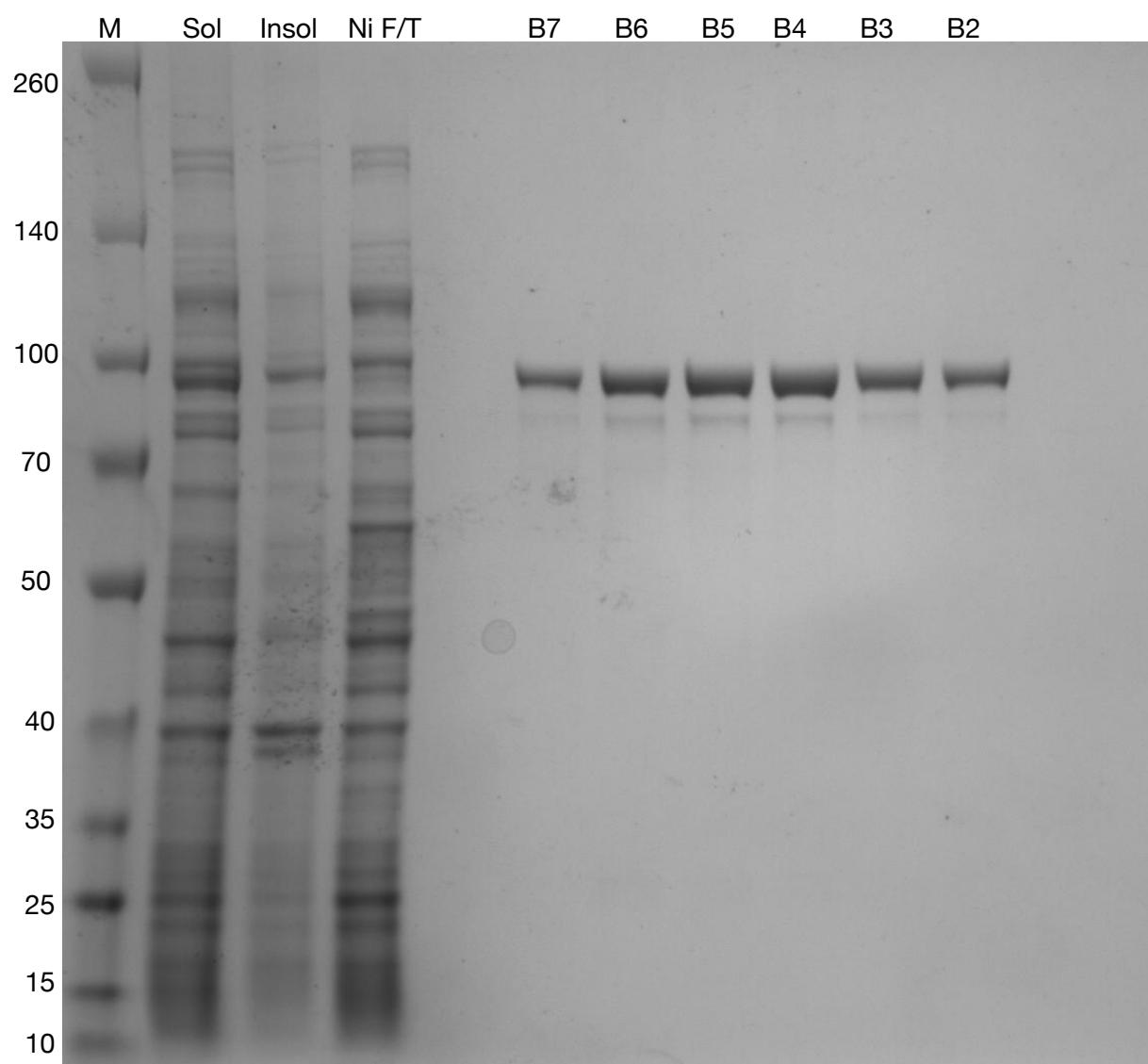
**Figure 5-2 - Supplementary Figure 6 - SDS-PAGE analysis of the tnPAP protein purification.**

Sol: Soluble fraction of cell lysate. Insol: Insoluble fraction of cell lysate. Ni F/T: The flow through after loading the sample onto the nickel column. A2-A10: Fractions collected from the gel filtration column believed to be the purified tnPAP protein with an expected MW of 61.1 kDa.



**Figure 5-2 - Supplementary Figure 7 - SDS-PAGE analysis of the tnAK protein purification.**

Sol: Soluble fraction of cell lysate. Insol: Insoluble fraction of cell lysate. Ni F/T: The flow through after loading the sample onto the nickel column. C3-D12: Fractions collected from the gel filtration column believed to be the purified tnPPT protein with an expected MW of 27.6 kDa.



**Figure 5-2 - Supplementary Figure 8 - SDS-PAGE analysis of the tePPK protein purification.**

Sol: Soluble fraction of cell lysate. Insol: Insoluble fraction of cell lysate. Ni F/T: The flow through after loading the sample onto the nickel column. B7-B2: Fractions collected from the gel filtration column believed to be the purified tePPK protein with an expected MW of 87.5 kDa.

**Figure 5-1 - Supplementary Figure 1 - Differential equations for modelling**

**afEst2**

$$\frac{dc_{Ester}}{dt} = -r1$$

$$\frac{dc_{Acid}}{dt} = r1$$

**mpCAR**

$$\frac{dc_{Acid}}{dt} = -r2$$

$$\frac{dc_{ATP}}{dt} = -r2$$

$$\frac{dc_{NADPH}}{dt} = -r2$$

$$\frac{dc_{Aldehyde}}{dt} = r2$$

$$\frac{dc_{Ppi}}{dt} = r2$$

$$\frac{dc_{AMP}}{dt} = r2$$

$$\frac{dc_{NADP+}}{dt} = r2$$

**mpCAR-ttPPIase**

$$\frac{dc_{Acid}}{dt} = -r2$$

$$\frac{dc_{ATP}}{dt} = -r2$$

$$\frac{dc_{NADPH}}{dt} = -r2$$

$$\frac{dc_{Aldehyde}}{dt} = r2 - r11$$

$$\frac{dc_{Ppi}}{dt} = r2 - r7$$

$$\frac{dc_{AMP}}{dt} = r2$$

$$\frac{dc_{NADP+}}{dt} = r2$$

**mpCAR-ttPPIase-PTDH**

$$\frac{dc_{Acid}}{dt} = -r2$$

$$\frac{dc_{ATP}}{dt} = -r2$$

$$\frac{dc_{NADPH}}{dt} = -r2 + r5$$

$$\frac{dc_{Aldehyde}}{dt} = r2 - r11$$

$$\frac{dc_{Ppi}}{dt} = r2 - r7$$

$$\frac{dc_{AMP}}{dt} = r2$$

$$\frac{dc_{NADP+}}{dt} = r2 - r5$$

$$\frac{dc_{PO3}}{dt} = -r5$$

$$\frac{dc_{PO4}}{dt} = -r5$$

**mpCAR-ttPPIase-PTDH-tnAK-tnPAP**

$$\frac{dc_{Acid}}{dt} = -r2$$

$$\frac{dc_{ATP}}{dt} = -r2 + r9 - r10$$

$$\frac{dc_{NADPH}}{dt} = -r2 + r5$$

$$\frac{dc_{Aldehyde}}{dt} = r2 - r11$$

$$\frac{dc_{Ppi}}{dt} = r2 - r7$$

$$\frac{dc_{AMP}}{dt} = r2 - r8 + r9 - r10$$

$$\frac{dc_{NADP+}}{dt} = r2 - r5$$

$$\frac{dc_{PO3}}{dt} = -r5$$

$$\frac{dc_{PO4}}{dt} = r5$$

$$\frac{dc_{PolyP}}{dt} = -r8$$

$$\frac{dc_{ADP}}{dt} = r8 - r9 + r10$$

**apADH**

$$\frac{dc_{Aldehyde}}{dt} = -r3 + r4 - r11$$

$$\frac{dc_{Alcohol}}{dt} = r3 - r4$$

$$\frac{dc_{NADH}}{dt} = -r3 + r4$$

$$\frac{dc_{NAD+}}{dt} = r3 - r4$$

**apADH-PTDH**

$$\frac{dc_{Aldehyde}}{dt} = -r3 + r4 - r11$$

$$\frac{dc_{Alcohol}}{dt} = r3 - r4$$

$$\frac{dc_{NADH}}{dt} = -r3 + r4 + r6$$

$$\frac{dc_{NAD+}}{dt} = r3 - r4 - r6$$

$$\frac{dc_{PO3}}{dt} = -r6$$

$$\frac{dc_{PO4}}{dt} = +r6$$

**afEst2-mpCAR-ttPPIase-PTDH-tnAK-tnPAP**

$$\frac{dc_{Ester}}{dt} = -r1$$

$$\frac{dc_{Acid}}{dt} = r1 - r2$$

$$\frac{dc_{ATP}}{dt} = -r2 + r9 - r10$$

$$\frac{dc_{NADPH}}{dt} = -r2 + r5$$

$$\frac{dc_{Aldehyde}}{dt} = r2 - r11$$

$$\frac{dc_{Ppi}}{dt} = r2 - r7$$

$$\frac{dc_{AMP}}{dt} = r2 - r8 + r9 - r10$$

$$\frac{dc_{NADP+}}{dt} = r2 - r5$$

$$\frac{dc_{PO3}}{dt} = -r5$$

$$\frac{dc_{PO4}}{dt} = r5$$

$$\frac{dc_{PolyP}}{dt} = -r8$$

$$\frac{dc_{ADP}}{dt} = r8 - r9 + r10$$

**mpCAR-ttPPIase-PTDH-tnAK-tnPAP-apADH**

$$\frac{dc_{Acid}}{dt} = -r2$$

$$\frac{dc_{ATP}}{dt} = -r2 + r9 - r10$$

$$\frac{dc_{NADPH}}{dt} = -r2 + r5$$

$$\frac{dc_{Aldehyde}}{dt} = r2 - r11 - r3 + r4$$

$$\frac{dc_{Ppi}}{dt} = r2 - r7$$

$$\frac{dc_{AMP}}{dt} = r2 - r8 + r9 - r10$$

$$\frac{dc_{NADP+}}{dt} = r2 - r5$$

$$\frac{dc_{PO3}}{dt} = -r5 - r6$$

$$\frac{dc_{PO4}}{dt} = r5 + r6$$

$$\frac{dc_{PolyP}}{dt} = -r8$$

$$\frac{dc_{ADP}}{dt} = r8 - r9 + r10$$

$$\frac{dc_{Alcohol}}{dt} = r3 - r4$$

$$\frac{dc_{NADH}}{dt} = -r3 + r4 + r6$$

$$\frac{dc_{NAD+}}{dt} = r3 - r4 - r6$$

**afEst2-mpCAR-ttPPIase-PTDH-tnAK-tnPAP-apADH**

$$\frac{dc_{Ester}}{dt} = -r1$$

$$\frac{dc_{Acid}}{dt} = r1 - r2$$

$$\frac{dc_{ATP}}{dt} = -r2 + r9 - r10$$

$$\frac{dc_{NADPH}}{dt} = -r2 + r5$$

$$\frac{dc_{Aldehyde}}{dt} = r2 - r11 - r3 + r4$$

$$\frac{dc_{Ppi}}{dt} = r2 - r7$$

$$\frac{dc_{AMP}}{dt} = r2 - r8 + r9 - r10$$

$$\frac{dc_{NADP+}}{dt} = r2 - r5$$

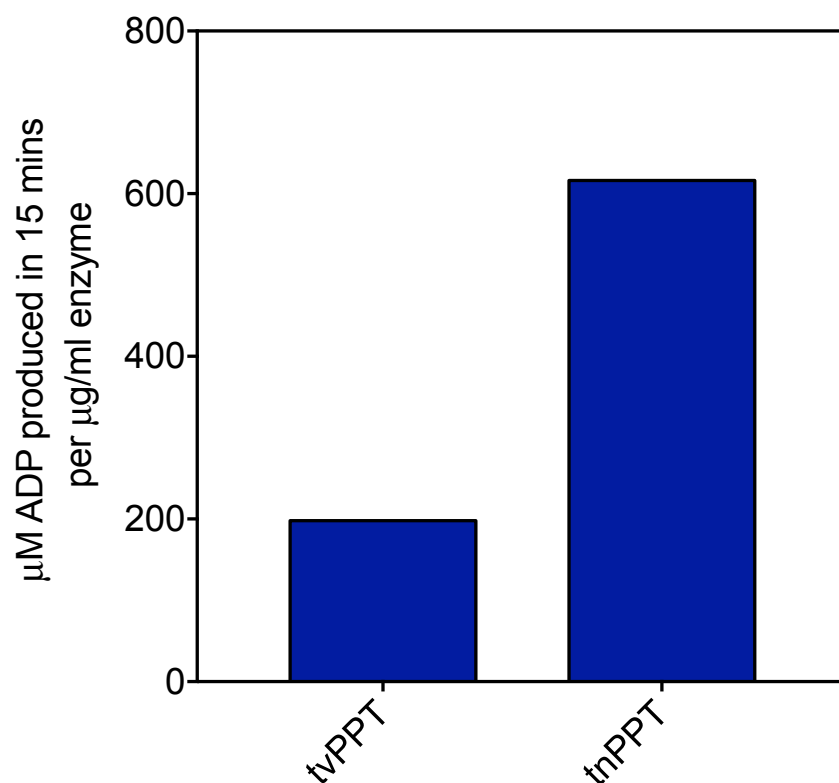
$$\frac{dc_{PO3}}{dt} = -r5 - r6$$

$$\frac{dc_{PO4}}{dt} = r5 + r6$$

$$\frac{dc_{PolyP}}{dt} = -r8$$

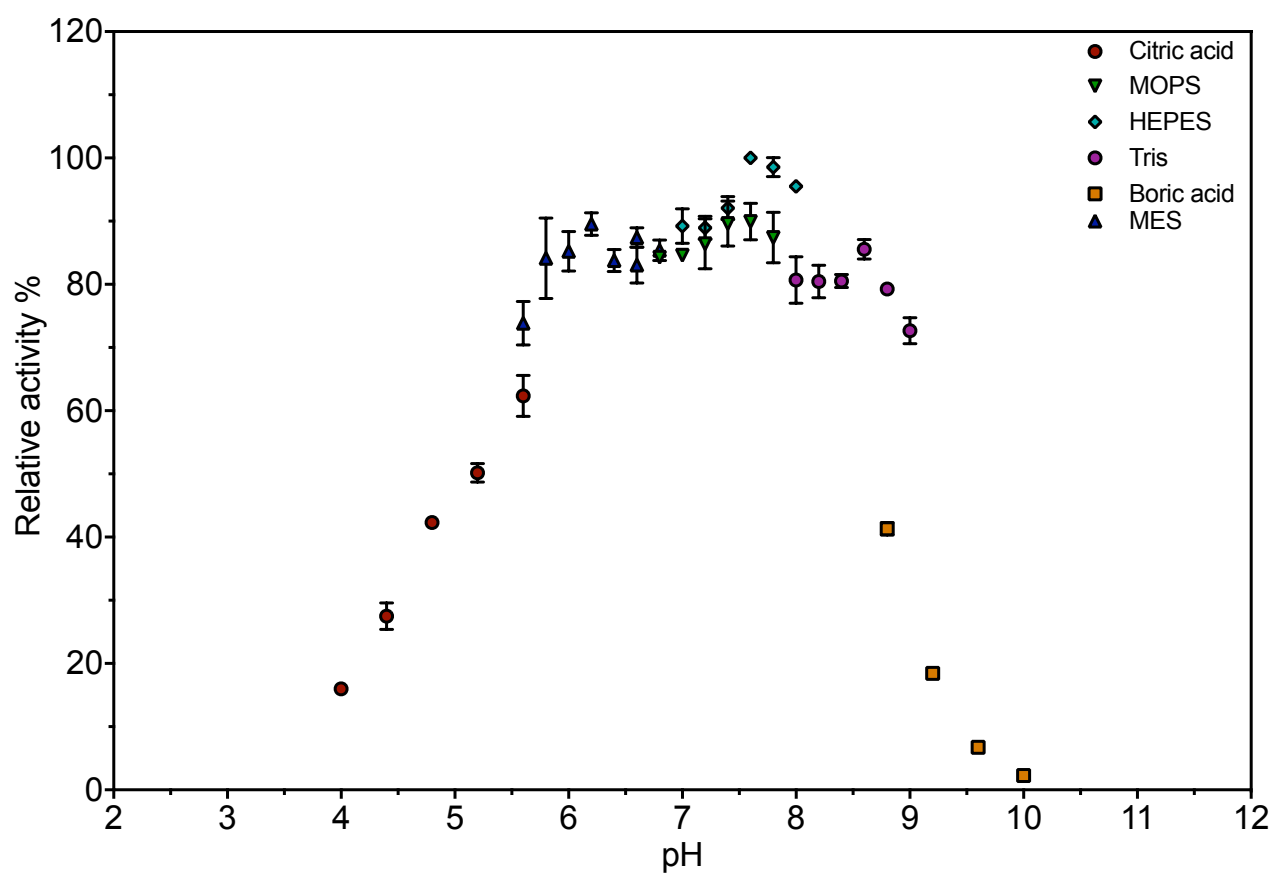
$$\frac{dc_{ADP}}{dt} = r8 - r9 + r10$$

$$\frac{dc_{Alcohol}}{dt} = r3 - r4$$



**Figure 5-2 - Supplementary Figure 9 – Preliminary data on two PAP enzymes.**

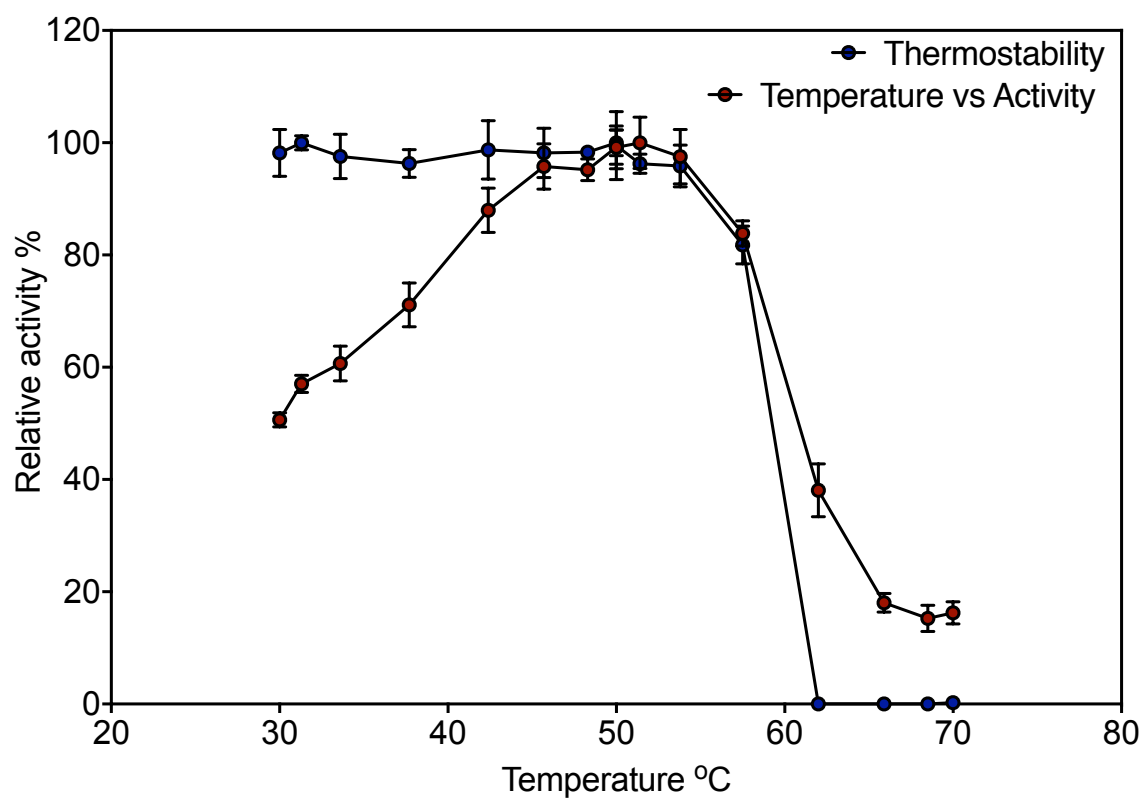
A preliminary experiment to determine presence of PAP activity, and which enzyme was most promising. ADP produced in 15 minutes divided by  $\mu\text{g} / \text{ml}$  enzyme used in the assay for comparison. The preliminary assay contained 5 mM  $\text{MgCl}_2$ , 1 mM AMP and 2.5 mM polyphosphate.



**Figure 5-3 - Supplementary Figure 1 – PTDH activity at pH**

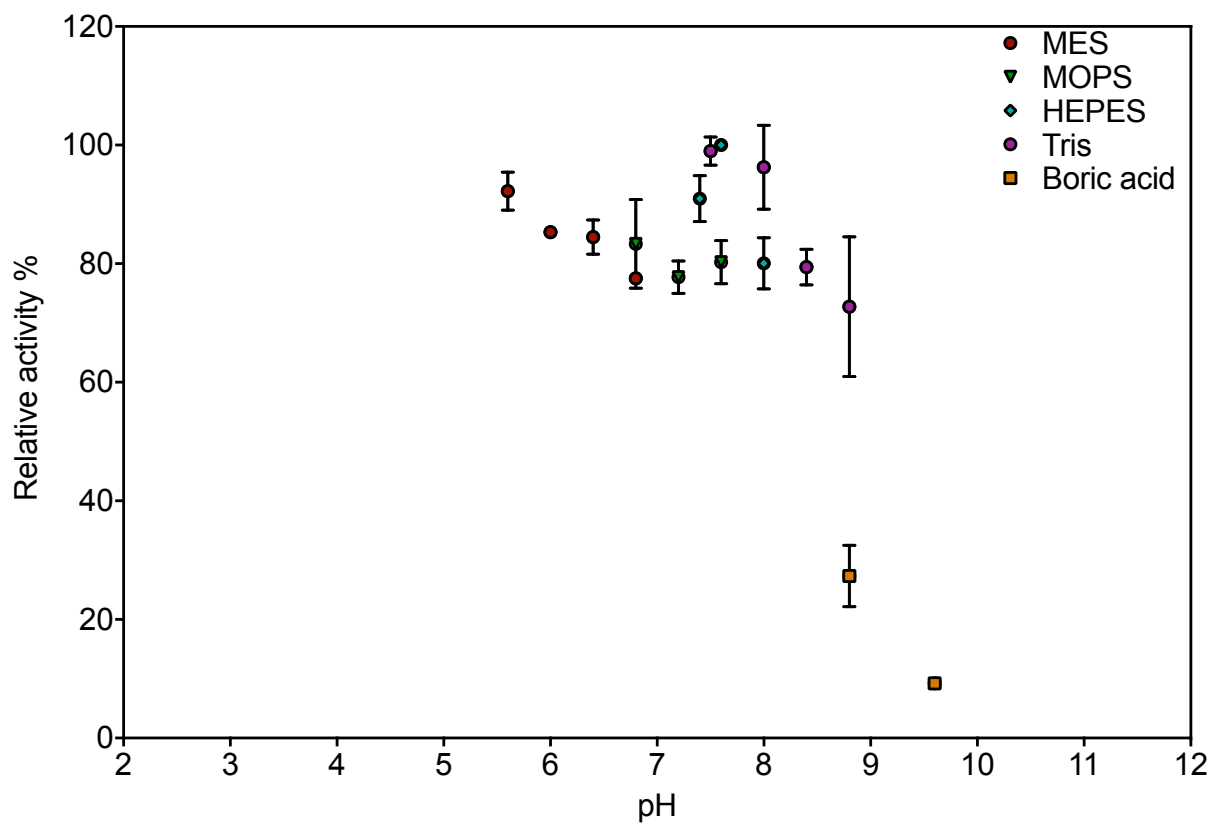
Activity of PTDH at various pH values, relative to the maximum activity detected. Various buffers were used to cover the range of pH values as indicated.





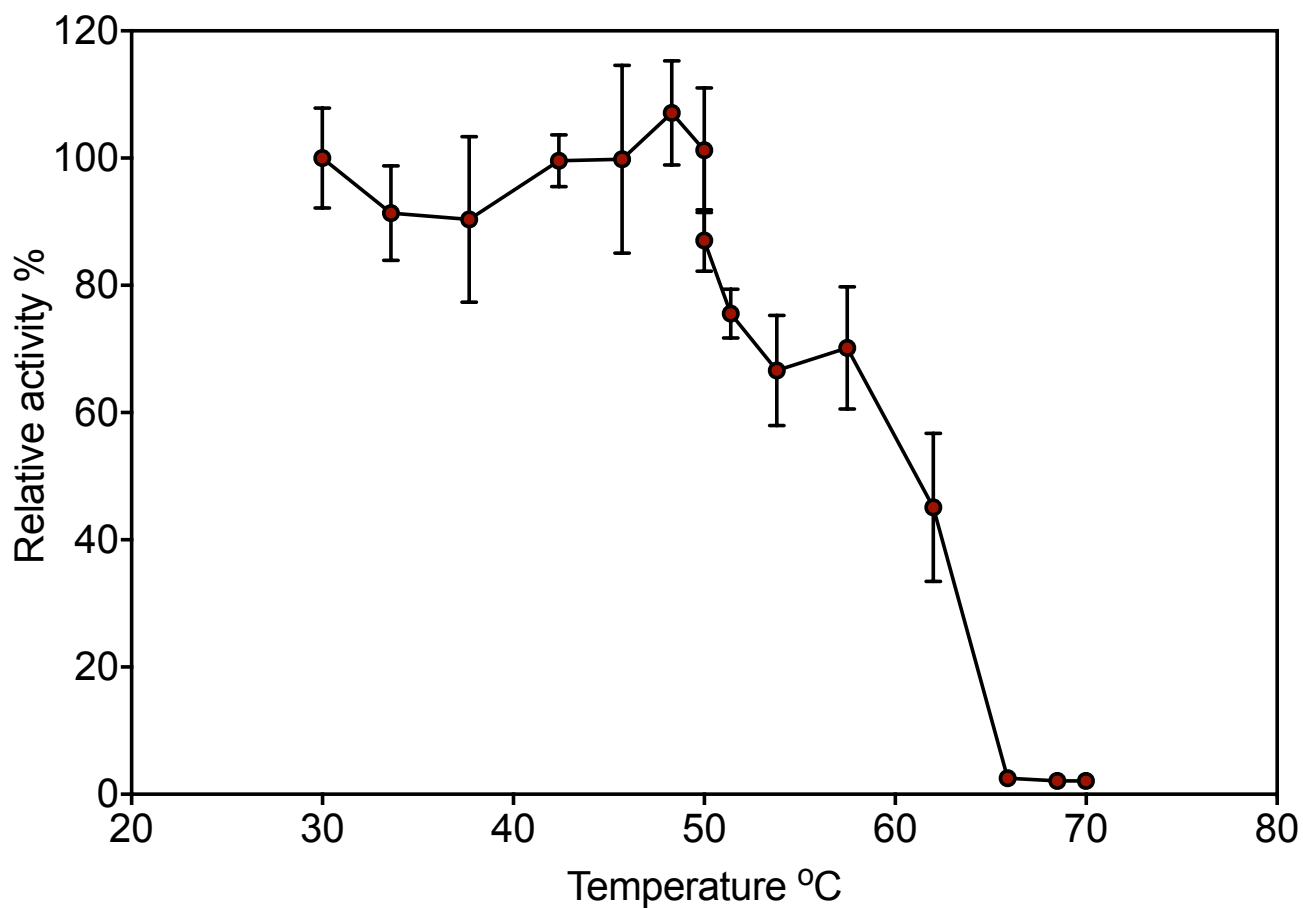
**Figure 5-3 - Supplementary Figure 2 – PTDH thermostability and temperature vs activity**

Red circles show PTDH activity at temperature, relative to the maximum activity at 51.4 °C. Blue circles show residual activity after incubating PTDH at various temperatures for 30 minutes, relative to a control kept on ice.



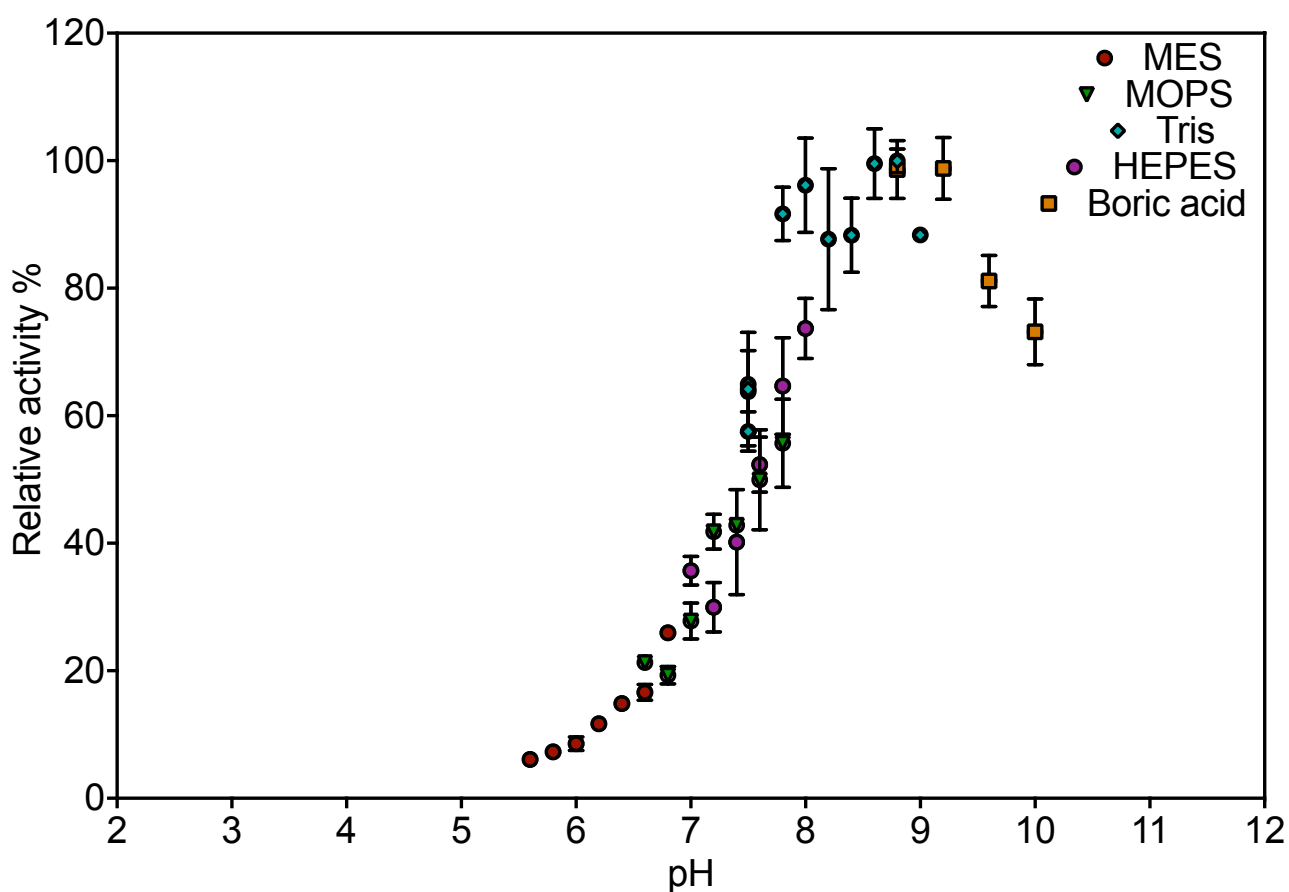
**Figure 5-3 - Supplementary Figure 3 – tnPAP activity at pH**

Activity of tnPAP at various pH values, relative to the maximum activity detected. Various buffers were used to cover the range of pH values as indicated.



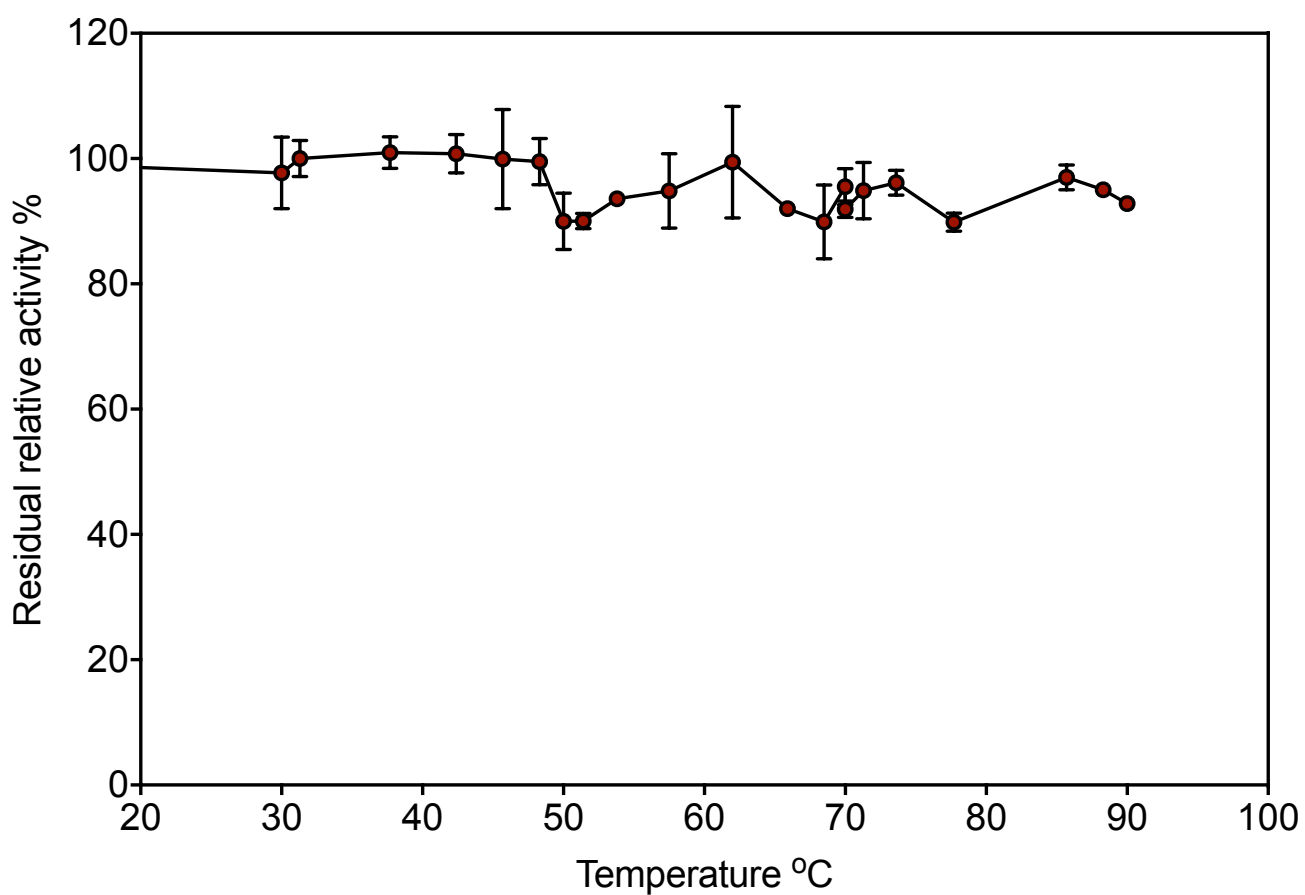
**Figure 5-3 - Supplementary Figure 4 – tnPAP thermostability**

Residual activity after incubating tnPAP at various temperatures for 30 minutes, relative to a control kept on ice.



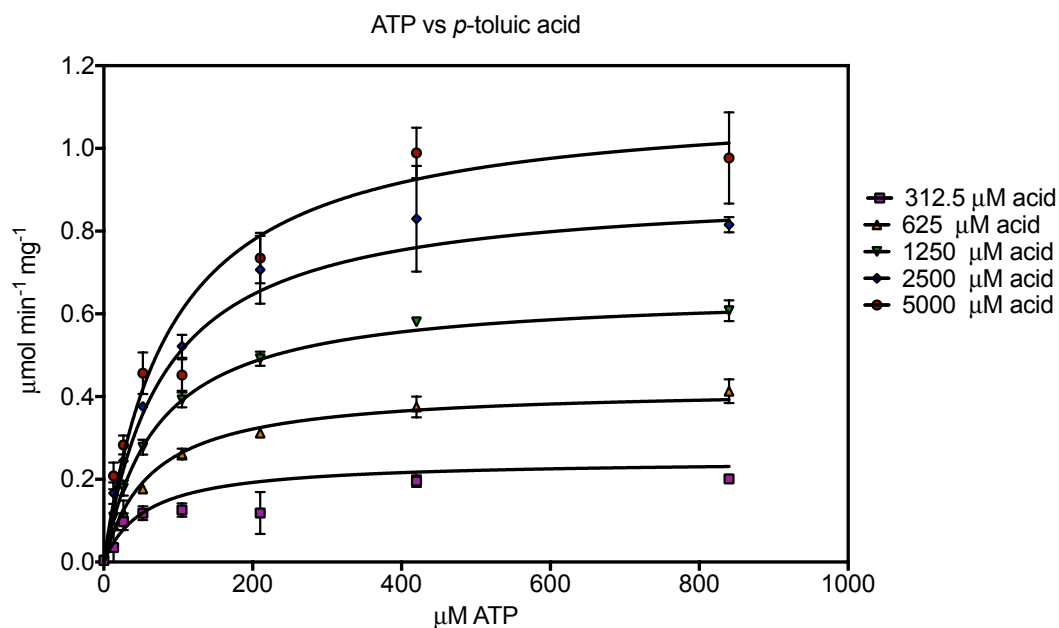
**Figure 5-3 - Supplementary Figure 5 – ttPPiase activity at pH**

Activity of ttPPiase at various pH values, relative to the maximum activity detected. Various buffers were used to cover the range of pH values as indicated.



**Figure 5-3 - Supplementary Figure 6 - ttPPiase thermostability**

Residual activity after incubating ttPPiase at various temperatures for 30 minutes, relative to a control kept on ice.



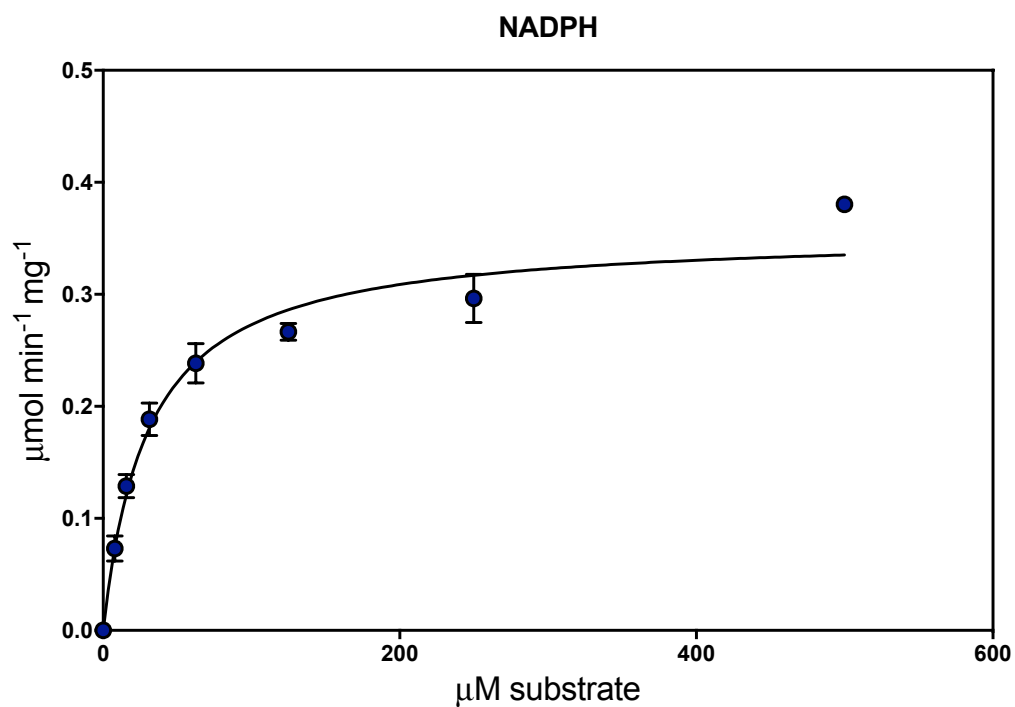
**Table 5-2 - Supplementary Figure 1 - The kinetics of mpCAR with varying concentrations of ATP and *p*-toluic acid.**

Data fit best to the equation for a steady-state sequential reaction.

**Table 5-2 - Supplementary Figure 2 – Parameters calculated fitting kinetics of mpCAR with varying concentrations of ATP and *p*-toluic acid to a sequential steady state equation.**

$$v = V_{max} \cdot \frac{[A] \cdot [B]}{(K_I^A \cdot K_M^B) + (K_M^A \cdot [B]) + (K_M^B \cdot [A]) + ([A] \cdot [B])}$$

Best-fit values	
$V_{MAX}$ (μmol / min / mg)	1.5
$K_I$ ATP (μM)	50
$K_M$ Acid (μM)	1500
$K_M$ ATP (μM)	100
Std. Error	
$V_{MAX}$ (μmol / min / mg)	0.07
$K_I$ ATP (μM)	20
$K_M$ Acid (μM)	200
$K_M$ ATP (μM)	10

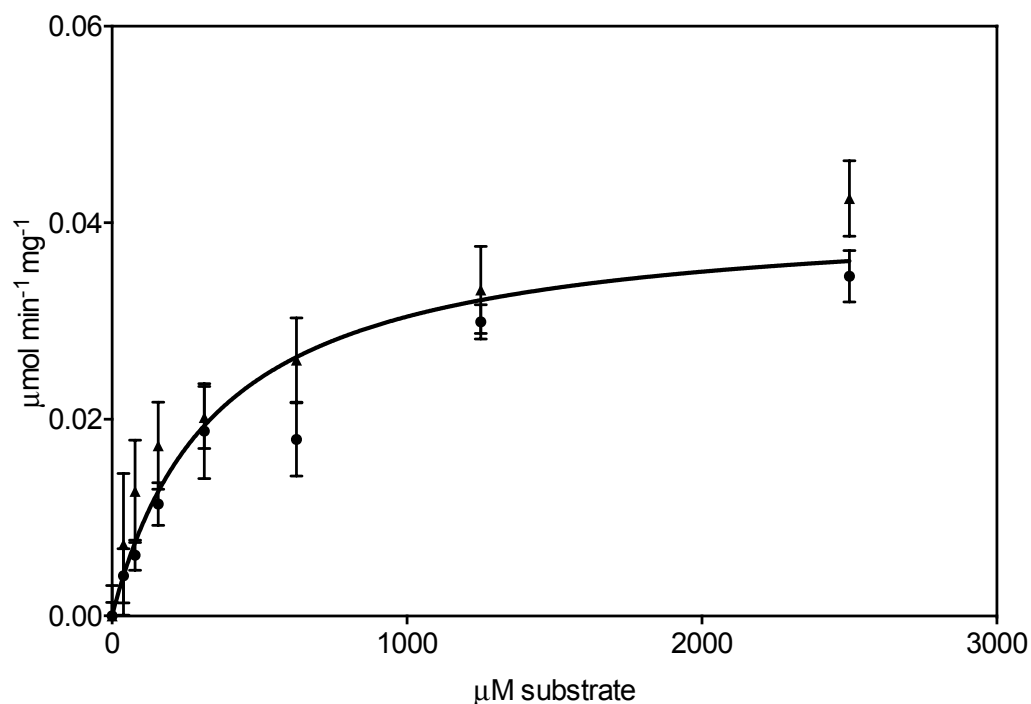


**Table 5-2 - Supplementary Figure 3 – The kinetics of mpCAR with varying concentrations NADPH.**

Data have been fit to the Michaelis-Menten equation.

**Table 5-2 - Supplementary Figure 4  
-Parameters calculated fitting  
kinetics of mpCAR with varying  
concentrations of NADPH to the  
Michaelis-Menten equation.**

<b>Best-fit values</b>	
$V_{MAX}$ (μmol / min / mg)	0.35
$K_M$ NADPH (μM)	30
<b>Std. Error</b>	
$V_{MAX}$ (μmol / min / mg)	0.01
$K_M$ NADPH (μM)	4



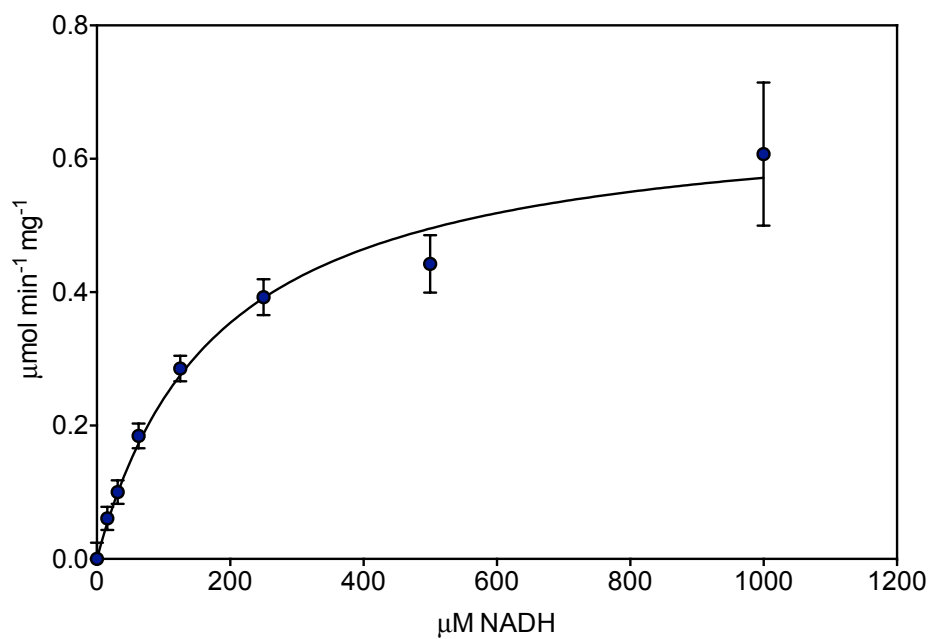
**Table 5-2 - Supplementary Figure 5 - The kinetics of apADH in the reductive direction at 30 °C, pH 7.5 with varying concentrations of *p*-tolualdehyde.**

Two experiments were carried out shown as triangles and circles, with the data combined for fitting to the Michaelis-Menten equation.

**Table 5-2 - Supplementary Figure 6 - Parameters calculated fitting kinetics of apADH in the reductive direction at 30 °C, pH 7.5, with varying concentrations of *p*-tolualdehyde to the Michaelis-Menten equation.**

<b>Best-fit values</b>	
$V_{MAX}$ (μmol / min / mg)	0.041
$K_M$ <i>p</i> -tolualdehyde (μM)	350
<b>Std. Error</b>	
$V_{MAX}$ (μmol / min / mg)	0.002
$K_M$ <i>p</i> -tolualdehyde (μM)	60





**Table 5-2 - Supplementary Figure 7 - The kinetics of apADH in the reductive direction at 70 °C, pH 7.5 with varying concentrations of NADH.**

Data have been fit to the Michaelis-Menten equation.

**Table 5-2 - Supplementary Figure 8 - Parameters calculated fitting kinetics of apADH in the reductive direction at 70 °C, pH 7.5, with varying concentrations of NADH, to the Michaelis-Menten equation.**

Only  $K_M$  used in this work.

**Best-fit values**

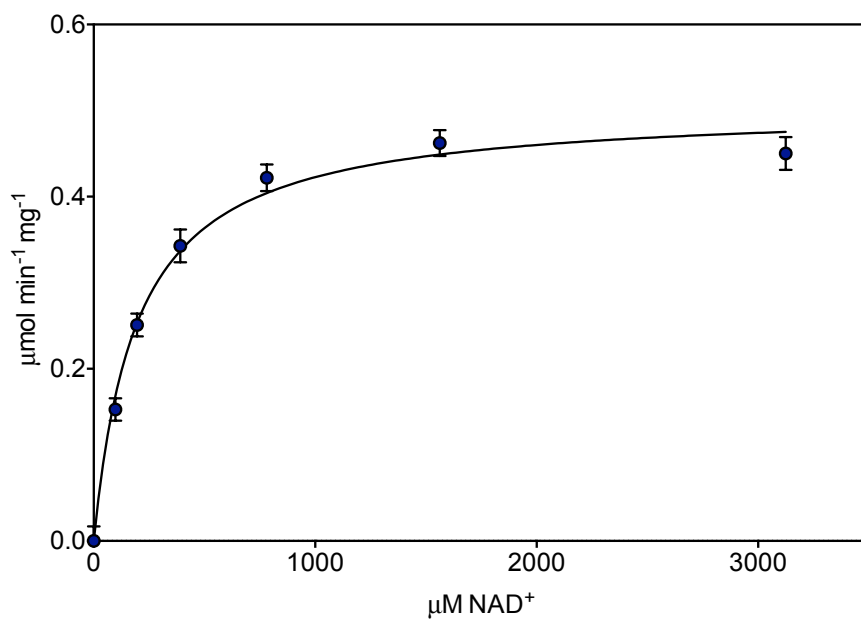
$V_{MAX}$ (μmol / min / mg)	0.68
-----------------------------	------

$K_M$ NADH (μM)	180
-----------------	-----

**Std. Error**

$V_{MAX}$ (μmol / min / mg)	0.04
-----------------------------	------

$K_M$ NADH (μM)	30
-----------------	----



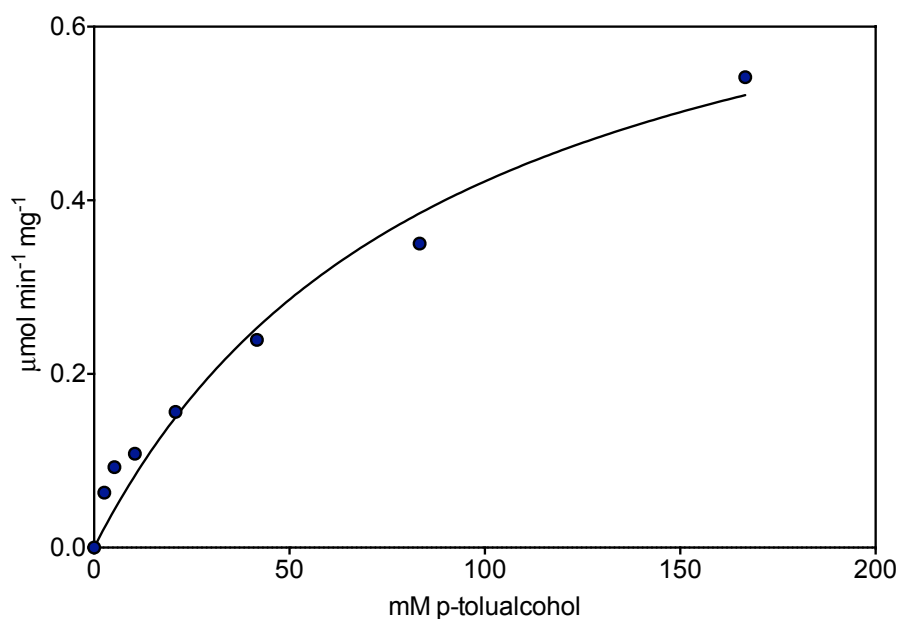
**Table 5-2 - Supplementary Figure 9 - The kinetics of apADH in the oxidative direction at 70 °C, pH 7.5 with varying concentrations of NAD<sup>+</sup>.**

Data have been fit to the Michaelis-Menten equation.

**Table 5-2 - Supplementary Figure 10 - Parameters calculated fitting kinetics of apADH in the oxidative direction at 70 °C, pH 7.5, with varying concentrations of NAD<sup>+</sup>, to the Michaelis-Menten equation.**

Only  $K_M$  used in this work.

<b>Best-fit values</b>	
$V_{MAX}$ ( $\mu\text{mol} / \text{min} / \text{mg}$ )	0.50
$K_M \text{ NAD}^+$ ( $\mu\text{M}$ )	195
<b>Std. Error</b>	
$V_{MAX}$ ( $\mu\text{mol} / \text{min} / \text{mg}$ )	0.01
$K_M \text{ NAD}^+$ ( $\mu\text{M}$ )	16



**Table 5-2 - Supplementary Figure 11 - The kinetics of apADH in the oxidative direction at 70 °C, pH 7.5 with varying concentrations of *p*-tolyl alcohol.**

Data fit to the Michaelis-Menten equation

**Table 5-2 - Supplementary Figure 12 - Parameters calculated fitting kinetics of apADH in the oxidative direction at 70 °C, pH 7.5, with varying concentrations of NAD<sup>+</sup>, to the Michaelis-Menten equation.**

As substrate concentration could not be taken high enough to calculate an accurate  $K_M$  or  $V_{MAX}$  this data was not used. An approximate  $K_M$  of 100 mM was used and  $k_{cat}$  in the oxidative direction assumed to be approximately equal to the forward.

**Best-fit values**

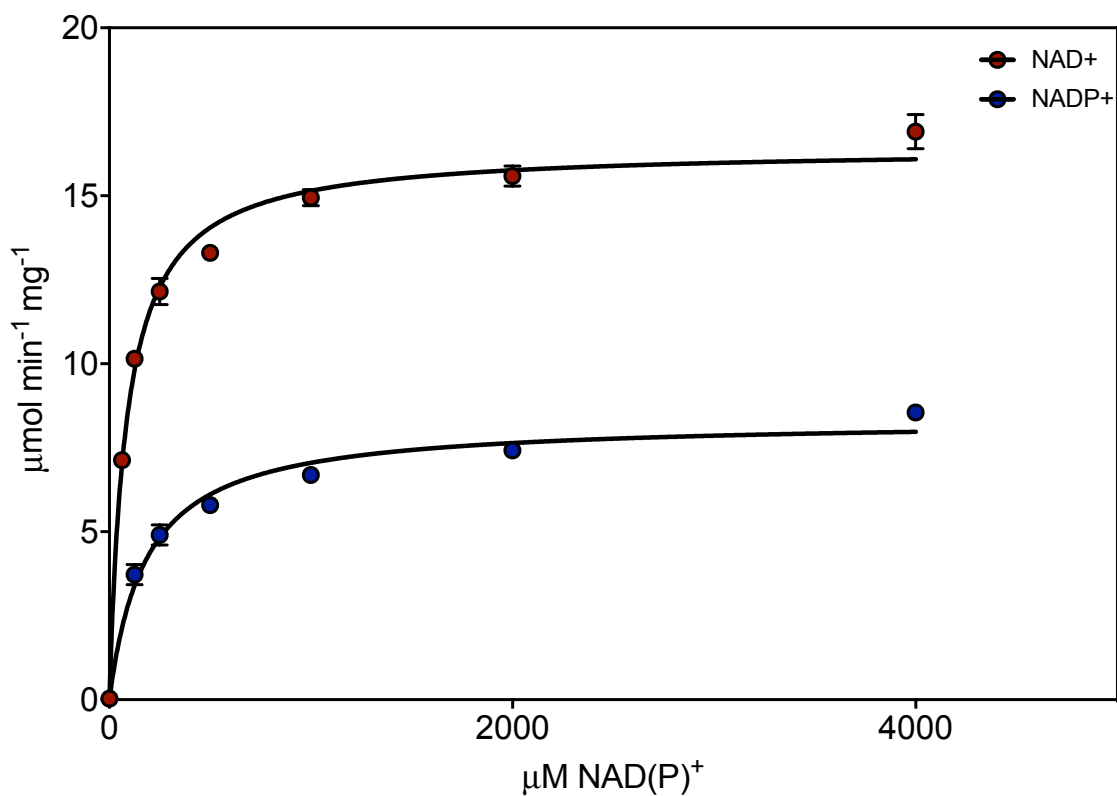
$V_{MAX}$ (μmol / min / mg)	0.80
-----------------------------	------

$K_M$ <i>p</i> -tolyl alcohol (mM)	90
------------------------------------	----

**Std. Error**

$V_{MAX}$ (μmol / min / mg)	0.06
-----------------------------	------

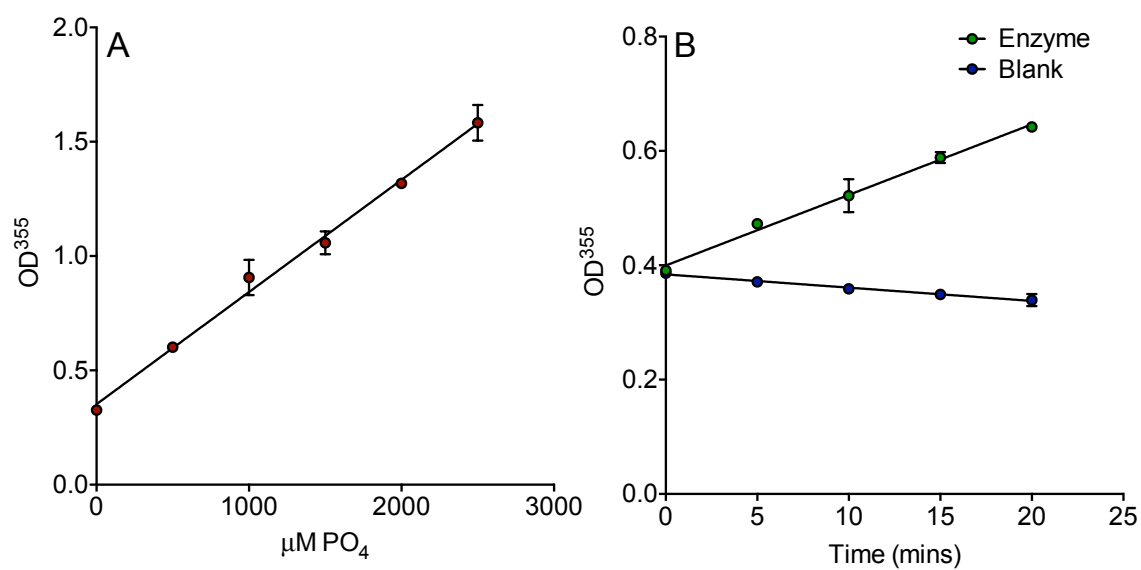
$K_M$ <i>p</i> -tolyl alcohol (mM)	15
------------------------------------	----



**Table 5-2 - Supplementary Figure 13 – The kinetics of PTDH with varying concentrations of NADP<sup>+</sup> and NAD<sup>+</sup>.**

Data fit to the Michaelis-Menten equation

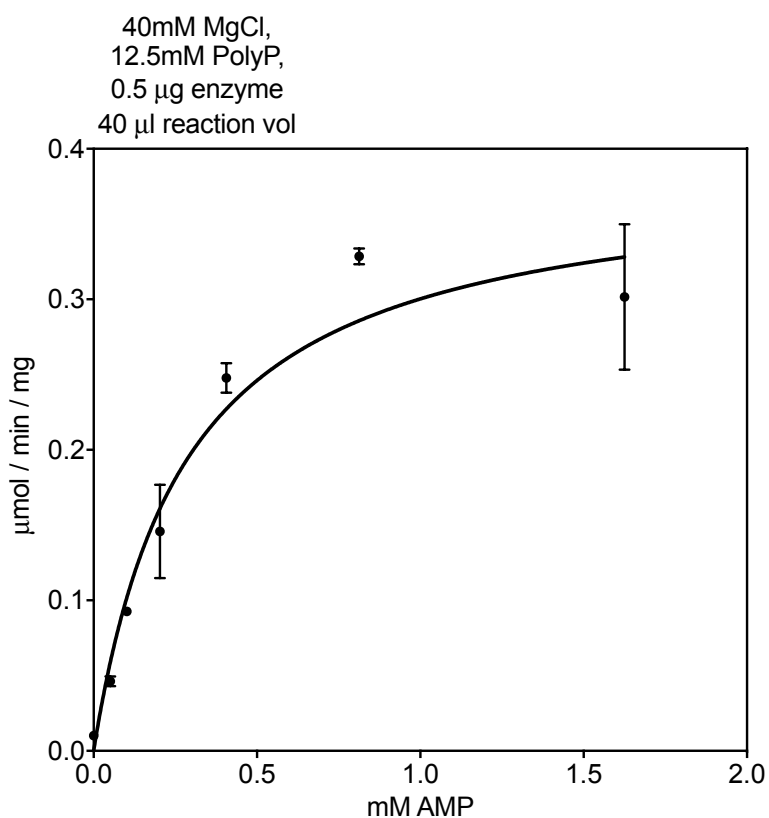
Table 5-2 - Supplementary Figure 14 - Parameters calculated fitting kinetics of PTDH with varying concentrations of NADP <sup>+</sup> and NAD <sup>+</sup> to the Michaelis-Menten equation.		
Best-fit values	NADP <sup>+</sup>	NAD <sup>+</sup>
$V_{MAX}$ (μmol / min / mg)	8.3	16.4
$K_M$ NAD(P) <sup>+</sup> (mM)	180	85
Std. Error		
$V_{MAX}$ (μmol / min / mg)	0.2	0.2
$K_M$ <i>p</i> -tolyl alcohol (mM)	20	5



**Table 5-2 - Supplementary Figure 15 – Determining  $k_{cat}$  for ttPPIase**

A: A standard curve of phosphate concentration vs OD 335 nm was determined for the assay.

B: The rate of phosphate production was determined (green circles), along with a blank rate (blue circles). Rate of blank subtracted phosphate production was calculated from the standard curve, from which  $k_{cat}$  was calculated.

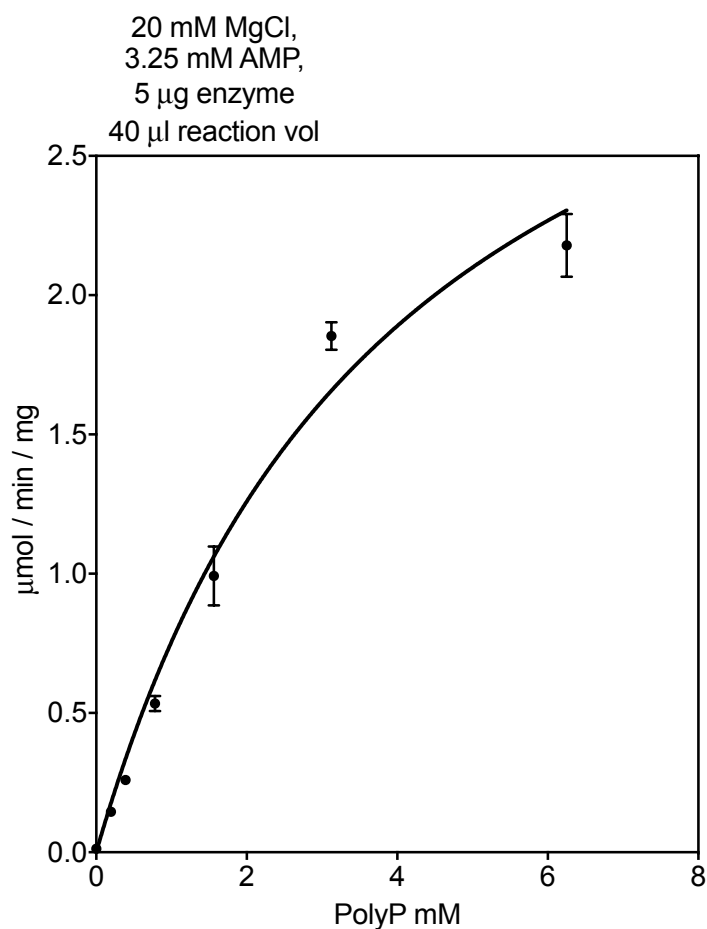


**Table 5-2 - Supplementary Figure 16 – The kinetics of tnPAP with varying concentrations of AMP**

Data fit to the Michaelis-Menten equation

**Table 5-2 - Supplementary Figure 17 - Parameters calculated fitting kinetics of tnPAP with varying concentrations of AMP to the Michaelis-Menten equation.**

Best-fit values	
$V_{MAX}$ (µmol / min / mg)	0.39
$K_M$ AMP (mM)	0.28
Std. Error	
$V_{MAX}$ (µmol / min / mg)	0.03
$K_M$ AMP (mM)	0.05

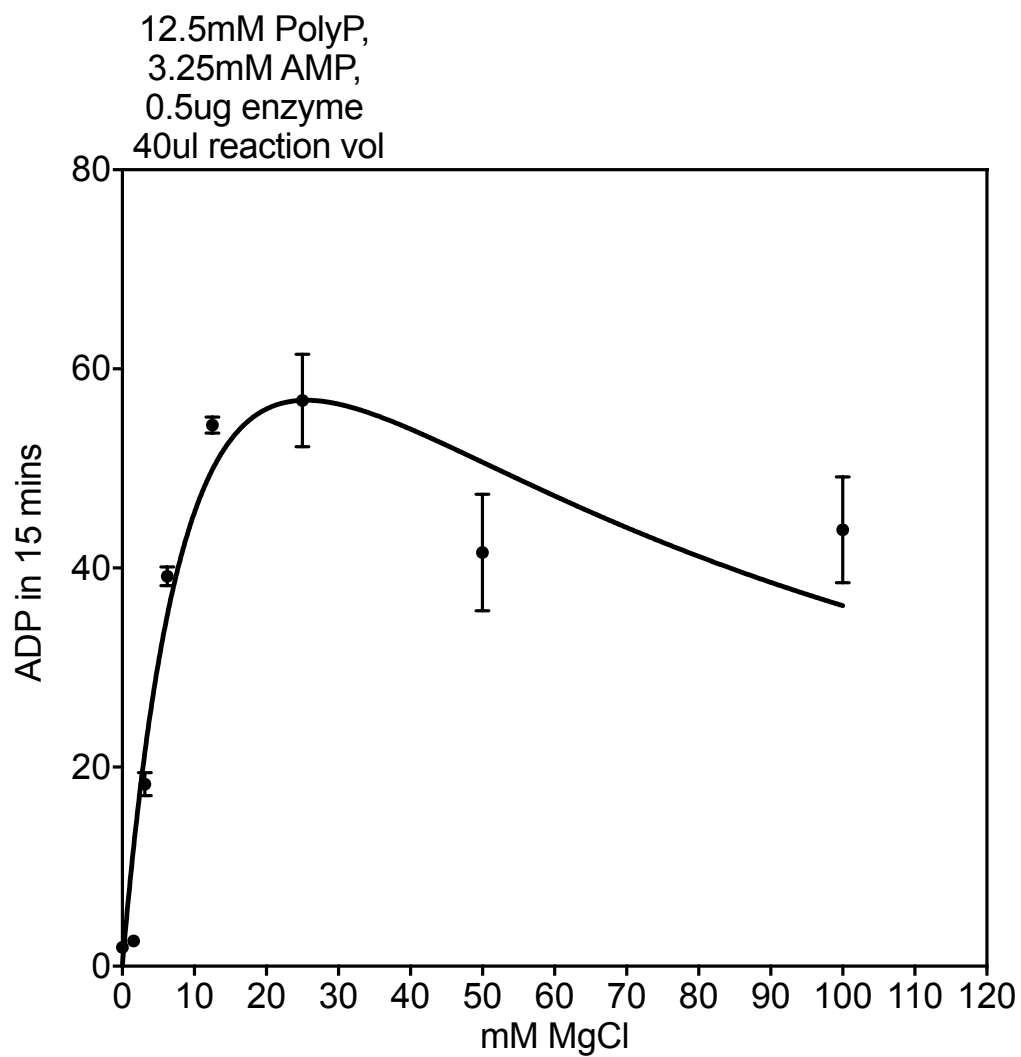


**Table 5-2 - Supplementary Figure 18 - The kinetics of tnPAP with varying concentrations of Polyphosphate (PolyP)**

Data fit to the Michaelis-Menten equation

**Table 5-2 - Supplementary Figure 19 - Parameters calculated fitting kinetics of tnPAP with varying concentrations of polyphosphate to the Michaelis-Menten equation.**

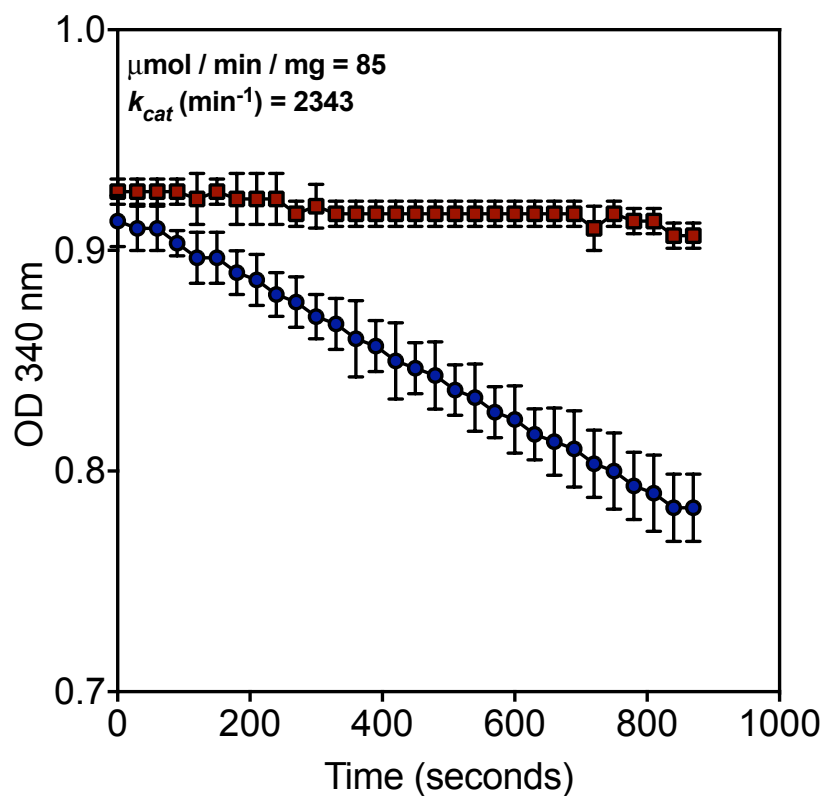
Best-fit values	
Vmax	3.781
Km	4.004
Std. Error	
Vmax	0.3454
Km	0.6808



**Table 5-2 - Supplementary Figure 20 - The kinetics of tnPAP with varying concentrations of MgCl**

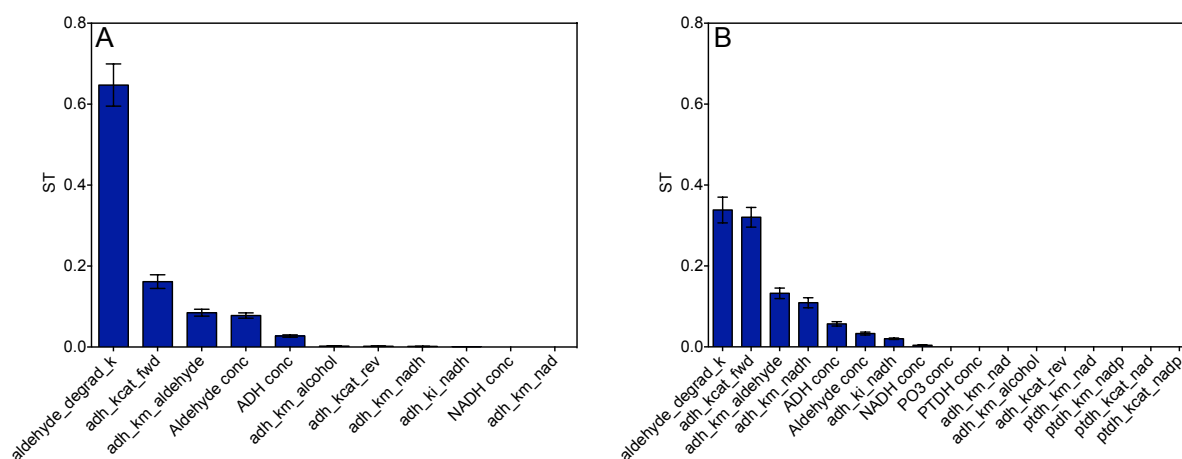
Data fit to the Michaelis-Menten equation with substrate inhibition





**Table 5-2 - Supplementary Figure 21 – tnAK – CAR coupled assay to estimate  $k_{cat}$**

tnAK was coupled to a CAR enzyme in order to estimate its  $k_{cat}$  in the ADP to ATP direction. The rate obtained was significantly slower than previously reported. The  $k_{cat}$  for the reverse reaction was adjusted relative. Red circles show the blank rate, blue squares show the rate with tnAK.



**Figure 5-6 - Supplementary Figure 1 – Sensitivity analysis of apADH and apADH-PTDH modelled reactions.**

The total sensitivity indices (ST) are shown which take into account 1st order and all other interactions. Sensitivity is in reference to the uncertainty in the final *p*-tolyl alcohol concentration. Error bars show the 95 % confidence intervals. The sum of all sensitivity indices' should equal 1.

A: Sensitivity analysis of apADH only reaction, figure 6A.

B: Sensitivity analysis of the apADH-PTDH reaction, figure 6B.

# **Chapter 6 - *Thermus thermophilus* as a host organism for whole-cell biocatalysis**

## 6.1 Introduction

### 6.1.1 *Thermus thermophilus* as a host organism for whole-cell biocatalysis

#### 6.1.1.1 Whole-cell biocatalysis using a thermophile as a host organism

Whole-cell biocatalysis offers a number of attractive benefits such as cofactor regeneration from the host metabolism, improved stability of enzymes inside the cell and no need for enzyme isolation. It can be particularly attractive for multi-step processes where multiple enzymes can be expressed in a single host organism, although these reactions are challenging to optimize (Ladkau, Schmid and Bühler, 2014).

Whole-cell biocatalysis using thermophiles as host organisms is a largely undeveloped area, with the potential to allow processes to be run at elevated temperatures (Taylor *et al.*, 2011). Elevated temperatures give a lower risk of contamination, potential for the recovery of volatile products, allows higher mass transfer rates with improved solubility of some substrates and can allow integration with chemical processes that require such temperatures (Frock and Kelly, 2012).

#### 6.1.1.2 *T. thermophilus* as an attractive host organism for whole-cell biocatalysis

*T. thermophilus* makes an excellent candidate for development as a host organism for thermophilic whole-cell biocatalysis. *T. thermophilus* is established as a model organism for the study of life at high temperatures (Cava, Hidalgo and Berenguer, 2009), has fast growth rates on complex media (Koyama *et al.*, 1986), and is naturally competent allowing easy transformation (Averhoff, 2009). This ease of transformation has allowed the development of a number of genetic tools (Averhoff, 2006). *T. thermophilus* is a thermophilic, Gram-negative, bacterium with an optimum growth temperature of 70 °C (Cava, Hidalgo and Berenguer, 2009), although it can grow between 50 and 85 °C (Sakaki and Oshima, 1975).

*T. thermophilus* HB27 is the strain most commonly investigated, as it demonstrates a higher transformation efficiency than other strains such as HB8, and maintains less

native plasmids. This difference in transformation efficiency is thought to be related to a less efficient modification-restriction barrier (Cava, Hidalgo and Berenguer, 2009). The genome sequence is also available for this strain (Henne *et al.*, 2004) and a structural genomics study has been carried out (Yokoyama *et al.*, 2000).

#### **6.1.1.3 *T. thermophilus* in the tree of life**

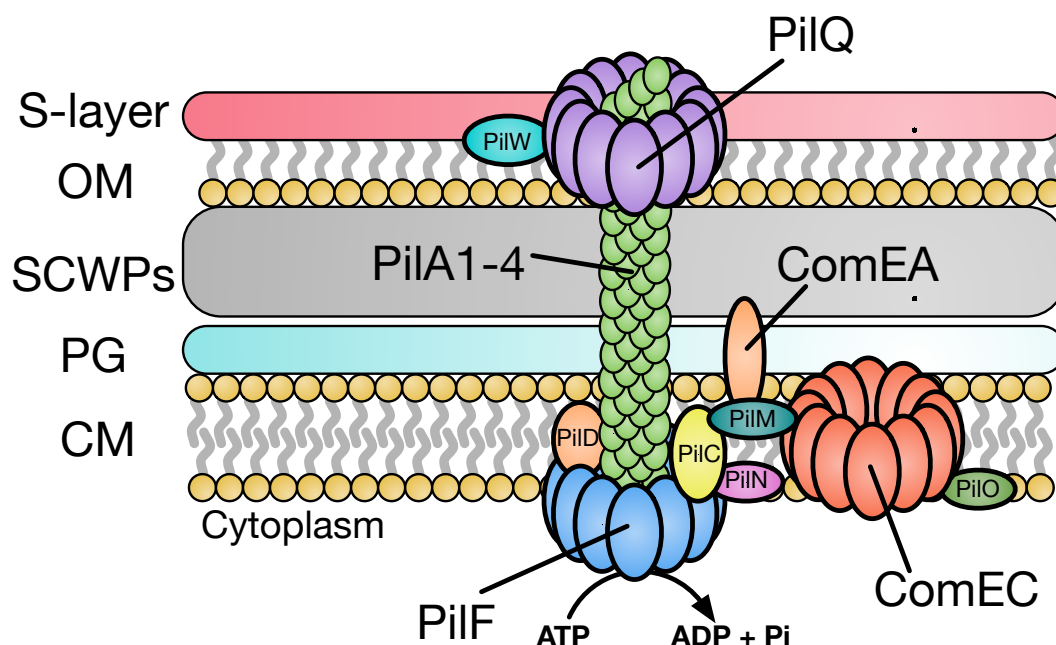
*T. thermophilus* belongs to the phylum Thermi (Langille *et al.*, 2013) (also known as Deinococcus-Thermus) which contains two distinct orders: the Deinococcales containing several species which are resistant to radiation, and the Thermales which contain many genera resistant to heat. *T. thermophilus* belongs to this latter order. While members of these two orders have evolved from a common ancestor and are both adapted to life in extreme environments, their mechanisms of resistance are thought to be largely independent (Omelchenko *et al.*, 2005).

#### **6.1.1.4 Natural competence by *T. thermophilus***

*T. thermophilus* is one of the few bacteria to possess natural competence over all stages of its growth and not to require any sort of stimulus to induce competence. This is due to the presence of a macromolecular DNA translocator assembly which mediates the entry of foreign DNA past the cell envelope into the cell (Averhoff, 2009). The ease with which *T. thermophilus* can be transformed has played a major role in its development as a model thermophilic organism (Cava, Hidalgo and Berenguer, 2009).

A model for this DNA translocation process has been developed as detailed in Figure 6-5 (Averhoff, 2009). Initially DNA is thought to bind to unknown DNA binding proteins on the cell surface or to proteins associated with a DNA translocator structure made up of pilin-like proteins. DNA is transported through or along a DNA transformation shaft made up of pilin PilA4 and some other minor pilin-like proteins such as PilA1, PilA2 and PilA3. PilQ proteins are thought to form a ring like structure and guide DNA translocator through the cell envelope. PilF, a putative DNA traffic NTPase, is thought to power DNA transport through the ring structure, and is possibly involved in the retraction of the DNA translocator. Once in the periplasmic space it is suggested that the DNA binds to the putative binding protein ComEA, which delivers the DNA to the inner membrane DNA transport complex. DNA is

thought to be transported through an inner membrane channel generated by ComEC, an inner membrane protein (Averhoff, 2004, 2006).



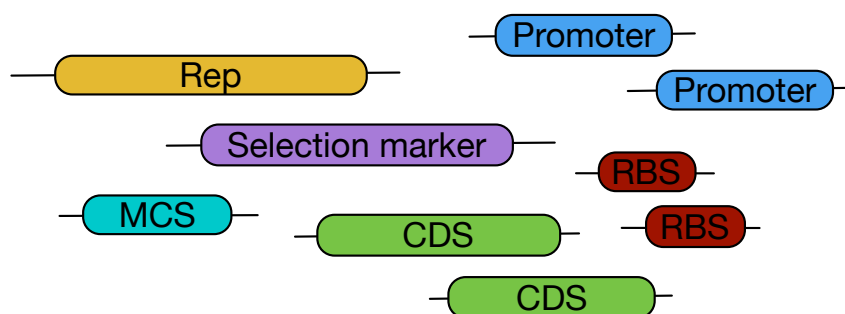
**Figure 6-1 – A model of the DNA transporter of *T. thermophilus***

DNA is transported along the DNA transformation shaft made up primarily of PilA4 and other pilin like proteins PilA1 to A3, guided through the outer membrane by the secretin ring (PilQ). Retraction of the DNA transformation shaft is powered by an ATPase PilF, which mediates transport into the periplasmic space. DNA then binds to a receptor protein ComEA where it is then delivered to an inner membrane channel thought to be made up by ComEC. Other proteins PilM, PilN, PilO and PilW are associated with the DNA transporter complex. OM: Outer membrane, SCWPs: secondary cell wall polymers, PG: peptidoglycan, CM: cytoplasmic membrane. The figure is adapted and redrawn from (Averhoff, 2009).

#### **6.1.1.5 Development of *T. thermophilus* as a host organism for whole-cell biocatalysis**

In order to develop and use *T. thermophilus* as a host organism for whole-cell biocatalysis, well defined synthetic biology parts are crucial (Rollié, Mangold and Sundmacher, 2012). While a number of vectors have been developed for *T.*

*thermophilus*, often these possess limited multiple cloning sites (MCSs) or large areas of apparently unrequired DNA. By identifying established vector parts, these can be combined for the creation of more efficient, well-designed vectors which facilitate standardized cloning approaches (Heinemann and Panke, 2006). Other well characterized parts such as promoters are also essential to facilitate a synthetic biology approach for the expression of biocatalytic pathways in *T. thermophilus* (Figure 6-2) (Rollié, Mangold and Sundmacher, 2012).



**Figure 6-2 – Well defined parts are essential for the development of *T. thermophilus* as a host organism for whole-cell biocatalysis.**

Defined and characterized parts that can be combined in a predictable way must be developed for *T. thermophilus* to be attractive for whole-cell biocatalysis. Regions of replication (Rep – yellow), selection markers (purple), multiple cloning sites (MCS – teal), coding sequences (CDS – green), ribosome binding sites (RBS – red) and promoters (blue) are all important for the construction of new pathways in *T. thermophilus*.

## 6.1.2 Vectors for *T. thermophilus*

### 6.1.2.1 Established vectors

A number of vectors have been established for *T. thermophilus* by altering cryptic plasmids with the addition of thermostable antibiotic resistance cassettes, MCSs, and regions of replication for *E. coli*. A good example and one of most commonly used or built upon vectors is pMK18. This consists of a fused region of replication from a cryptic plasmid of *Thermus* sp. ATCC27737, a thermostable kanamycin resistance cassette, and a replicative origin and MCS from pUC18 (de Grado and Casta, 1999). Another commercially available vector is pMKE2, derived from pMK18

with the inclusion of the Pnar promoter for inducible protein expression in *T. thermophilus* (Moreno *et al.*, 2003). A number of other vectors have been constructed in a similar fashion (Averhoff, 2006).

#### **6.1.2.2 Selectable markers at elevated temperatures**

Selectable markers are essential for the use of a vector in *T. thermophilus*, or for insertions or deletions on the *T. thermophilus* genome. The most commonly used is a thermostable kanamycin resistance cassette. The promoter for the S-layer protein, PslpA, is commonly used to drive the expression in the cassette, and allows function in both *T. thermophilus* and *E. coli*. A number of other selectable markers are also available which can also be used in conjunction with the kanamycin resistance marker (Table 6-1).



Table 6-1 – Selectable markers for <i>T. thermophilus</i>					
Gene	Name	NCBI Reference	Media requirements	Strain requirements	Description
<i>kat</i>	Kanamycin resistance (thermostable)	GI:363583404	20 - 40 µg / ml kanamycin	None	Antibiotic resistance. Kanamycin adenyl transferase has been engineered for improved thermostability. Expression from the PslpA promoter is common and allows resistance to kanamycin in both <i>T. thermophilus</i> , and <i>E. coli</i> (Matsumura and Aiba, 1985; Lasa <i>et al.</i> , 1992).
<i>ble</i>	Bleomycin resistance (thermostable)	GI:60594050	3 - 15 µg / ml bleomycin	None	Antibiotic resistance. Bleomycin-binding protein is a small, highly negatively charged protein which can bind positively charged antibiotic molecules such as bleomycin. It was been engineered by directed evolution for thermostability and can also function in <i>E. coli</i> when expressed from PslpA (Brouns <i>et al.</i> , 2005).

<i>hyg</i>	Hygromycin resistance (thermostable)	GI:213020041	40 µg / ml hygromycin	None	Antibiotic resistance. A thermostable <i>E. coli</i> hygromycin phosphotransferase was developed using directed evolution in <i>T. thermophilus</i> . It is stable up to 65 °C and has been used in both a plasmid and as a selection marker for genome integration (Nakamura <i>et al.</i> , 2005).
<i>pyrE</i>	Orotate phosphoribosyltransferase	GI:499487123	Minimal media containing 50 µg / ml uracil, or 200 µg / ml 5-FOA	<i>pyrE</i> knockout strain	A knockout of <i>pyrE</i> results in resistance against the bactericidal compound 5-fluoroorotic acid (5-FOA), and uracil auxotrophy. Can be used effectively for iterative knock-ins or knock-outs (Tamakoshi <i>et al.</i> , 1997, 1999).
<i>leuB</i>	3-isopropyl-malate dehydrogenase	GI:21262177	Minimal media containing 50 µg / ml leucine and isoleucine.	Auxotrophic strain	Auxotrophic strains can be complemented on a vector, or used as a marker for successful knock-ins. The use of minimal media with or without the relevant amino acid is required. The knockout of <i>leuB</i> , which codes for a 3-isopropyl-malate dehydrogenase, causes leucine auxotrophy and has been used in this way (Tamakoshi, Yamagishi and Oshima, 1995).

<i>mdh</i>	Malate dehydrogenase	GI:46196098	None	<i>mdh</i> knockout strain	Malate dehydrogenase is a key enzyme in the tricarboxylic acid cycle and <i>mdh</i> deficient cells show a small colony phenotype, even after 5 days of growth. This can allow easy identification of colonies that have been successfully transformed or have had <i>mdh</i> knocked out (Kayser and Kilbane, 2001).
<p>Selectable markers are essential for the design of new vectors for protein expression in <i>T. thermophilus</i>. However the use of standard markers is limited due to the requirement for thermostability. Three antibiotic resistance cassettes have been developed as well as a number of other markers that can be used to select for successful transformants. The most useful of these are detailed in the table.</p>					

### 6.1.2.3 Cryptic plasmids – regions of replication

*T. thermophilus* HB8 has three plasmids, pTT8, pTT27 and pVV8, which can be used for the construction of new synthetic plasmids (Ohtani, Tomita and Itaya, 2013). In contrast *T. thermophilus* HB27 has only pTT27, and efforts to remove this plasmid have resulted in a plasmid free strain, better suited for industrial exploitation (Ohtani, Tomita and Itaya, 2016). In addition to these plasmids, a region of replication taken from a *Thermus* sp. ATCC27727 plasmid has been used for the construction of synthetic vectors such as pMK18 for *T. thermophilus* (de Grado and Casta, 1999). More recently a small multi-copy cryptic plasmid from *Thermus* sp. TK10 has been used as a replication region for the creation of a *T. thermophilus* vector, pNHK101 (Kobayashi *et al.*, 2005).

### 6.1.2.4 Region of replication from a cryptic plasmid of *Thermus* sp.

#### ATCC27737

The minimum replication region identified from the 16 kbp plasmid of *Thermus* sp. ATCC27737 is 1.8 kbp in size and encodes a RepA protein which is essential for replication of the plasmid. The N-terminal fragment of RepA binds an AT rich area of DNA surrounded by GC rich sequences in its own coding sequence. This is thought to be the origin of replication (De Grado, Lasa and Berenguer, 1998). This region of replication was used for the construction of pMK18 and pMKE2, and was used for the construction of vectors in this study.

### 6.1.2.5 Region of replication for the cryptic plasmid pTT8

The replication region from the pTT8 9.3 kbp cryptic plasmid of *T. thermophilus* HB8 was identified as a 1.6 kbp region. This region shows homology to the replication regions of ColE2 related plasmids (Aoki and Itoh, 2007). The region of replication contains the gene for a replication initiator protein Rep. pTT8 also contains genes for PasA and PasB proteins, which function as a stability system for pTT8 (Takayama *et al.*, 2004). The origin of replication recognized by Rep was identified as a region just downstream of the *rep* gene. Rep expression is regulated by anti-sense RNA just upstream of the *rep* gene (Aoki and Itoh, 2007). A high copy number mutant of pTT8 was obtained in which this region was removed and replaced with a fragment of DNA containing two promoter-like elements and an RBS

just upstream of the *rep* ORF. While the normal copy number of pTT8 is approximately eight, the mutant showed a copy number of 30 to 40, with very good stability (Takayama *et al.*, 2004).

#### **6.1.2.6 Region of replication from the pVV8 mega-plasmid**

The replication region from the pVV8 81.2 kbp mega-plasmid was identified by stepwise deletion to a region containing a single ORF TTHV001, RepV. RepV has been shown to bind to an inverted repeat in its own DNA sequence, triggering an unwinding of the DNA duplex at an AT rich region downstream of this site. This replication region is thought to show promise for co-existence with other plasmids (Ohtani, Tomita and Itaya, 2013).

#### **6.1.2.7 A cryptic plasmid from *Thermus* sp. TK10 as a replication region**

A small 1.5 kbp multi-copy cryptic plasmid from *Thermus* sp. TK10 has been used as a replication region for the construction of a *T. thermophilus* expression vector pNHK101. No significant similarities to other plasmids have been identified, and only four ORFs are found within it. Three of these ORFs show small similarities to several replication proteins from other plasmids. The vector showed a copy number of approximately 80 in *T. thermophilus*, which reduced to about 20 with the addition of a thermostable kanamycin resistance cassette. pNHK101 was unstable above 60 °C, but stability was improved with the addition of pTT8 at 70 °C. This vector shows promise for co-existence with other plasmids, but more work is needed to understand its instability and replication mechanism (Kobayashi *et al.*, 2005).

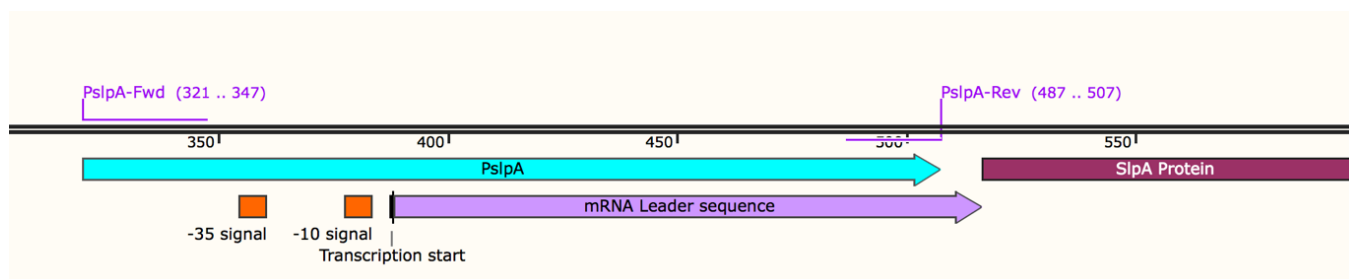
#### **6.1.2.8 Replication region of the pTT27 mega-plasmid**

pTT27 is a mega-plasmid of about 233 kbp found in both *T. thermophilus* HB8 and HB27, although there are number of differences between the plasmids of the two strains. A minimal region for replication was identified as a 2.3 kbp region, which includes a single gene for a replication initiation protein RepT, which binds to direct repeats downstream of its ORF. A region for the expression of subunits of a class I ribonucleotide reductase for deoxynucleoside triphosphate synthesis was also shown to be essential for plasmid maintenance. For plasmids designed from pTT27 to be used in *T. thermophilus* HB27, a strain which has had the natural pTT27 plasmid removed is likely to be necessary (Ohtani, Tomita and Itaya, 2016).

### 6.1.3 Promoters for *T. thermophilus*

#### 6.1.3.1 Constitutive promoter - PslpA

The promoter for the S-layer protein, PslpA (Figure 6-3), is a very strong constitutive promoter, however it is also tightly regulated. The promoter is negatively regulated by the SlrA and SlpM proteins. SlpA is also able to regulate its own translation by binding to a leader region of its own mRNA preventing translation initiation (Fernández-Herrero, Olabarría and Berenguer, 1997; Cava, Hidalgo and Berenguer, 2009). Fusions of PslpA to *lacZ* have shown regulation both transcriptionally and translationally to be dependent on the 5' UTR [14]. Possibly the removal of this region would result in a better synthetic promoter, although it has been speculated that the 5' UTR may have implications for mRNA stability [14]. The PslpA promoter is also able to function in *E. coli*, making this a very useful promoter for the expression of antibiotic resistance cassettes which can work in both *T. thermophilus* and *E. coli* (Moreno *et al.*, 2003).



**Figure 6-3 – A map of PslpA promoter and the beginning of the *slpA* gene.**

-35 and -10 promoter regions are shown as well as the transcription start site. Much of the region identified as the promoter is transcribed as an mRNA leader sequence. This region has been shown to be important in the regulation of PslpA. This figure was made with SnapGene.

#### 6.1.3.2 Constitutive promoter - PrmpB

The promoter for the RmpB 50S ribosomal protein, PrmpB, is a constitutive promoter that is weaker than PslpA, which offers a more moderate level of protein expression. It was identified as a useful promoter (personal communication with José Berenguer and Aurelio Hidalgo, Universidad Autónoma de Madrid, Spain).

#### **6.1.3.3 Nitrate and anoxic conditions inducible promoter - Pnar**

The promoter region of the respiratory nitrate reductase operon from *T. thermophilus* HB8, Pnar, allows inducible expression by the addition of nitrate and anoxic conditions (Moreno *et al.*, 2003). Two transcriptional regulators, DnrS and DnrT responding to anoxia and nitrate respectively, are found in the *nar* operon and are essential for the induction of Pnar (Moreno *et al.*, 2003). The *nar* operon is not normally found in *T. thermophilus* HB27, however it can be transferred from HB8 by conjugation (Cava *et al.*, 2004). Reasonably good levels of protein expression have been reported using the Pnar promoter (Hidalgo *et al.*, 2004; Moreno *et al.*, 2005), although it requires the use of anoxic conditions severely stunting growth.

#### **6.1.3.4 Constitutive promoter – Pnqo**

The *nqo* operon encodes a type I NADH dehydrogenase that is the main electron donor during aerobic growth. Its promoter Pnqo is constitutive but repressed by DnrT under anaerobic conditions and by the addition of nitrate (Cava *et al.*, 2007). This promoter offers weaker constitutive expression than PslpA.

#### **6.1.3.5 Heat inducible promoter - PdnaK**

The promoter for the *dnaK* gene in *T. thermophilus* is activated by heat shock, specifically transferring cultures from 70 °C to 85 °C. When this promoter was used in the expression of malate dehydrogenase as a reporter gene it was shown to give significantly high background levels, and an increase of ~38% on heat induction (Kayser *et al.*, 2001). In another study using  $\beta$ -galactosidase as a reporter the high background expression was confirmed, but a doubling of expression on heat shock was demonstrated (Park and Kilbane, 2004).

#### **6.1.3.6 Arginine and carbon source inducible promoters**

An arginine inducible promoter Parg, has been shown to be induced by the addition of 10 to 30 mM arginine. However only a maximum of a three-fold increase in expression was detected using this promoter (Park and Kilbane, 2004). Another promoter, Pscs-mdh, is induced by the addition of carbon sources such as malate, succinate, pyruvate, glutamine, glucose, yeast extract or peptone (Park and Kilbane, 2004).

#### **6.1.3.7 Potential cold shock promoter - Pcsp2**

The cold shock protein gene *csp2* shows a rapid and large increase in mRNA following a temperature downshift in *T. thermophilus* HB8. This is thought to be due to a temperature dependent secondary structure forming at 45 °C at the 5' UTR of the mRNA, assumed to prevent degradation of the mRNA at this temperature (Mega *et al.*, 2010). This 5' UTR or even the whole *csp2* promoter could provide an interesting approach for the inducible expression of proteins in *T. thermophilus*, upon dropping temperature to 45 °C, causing the cell growth to slow considerably. However the use of Pcsp2 for the expression of other proteins has never been tested.

#### **6.1.3.8 Identification of promoters from genomic DNA fragments**

Other bifunctional constitutive promoters have been developed by using a promoter-less kanamycin resistance gene on a promoter-probe vector, and ligating random chromosomal fragments from *T. thermophilus* or *T. flavus* into this vector, and screening for kanamycin resistant colonies. Transformants capable of surviving on the high levels of kanamycin were selected. Of those, two promoters termed J17 and vv12 were selected that also conferred resistance in *E. coli*, although the sequences are not available (Kayser *et al.*, 2001).

### **6.1.4 Ribosome binding sites for *T. thermophilus***

#### **6.1.4.1 Ribosome binding site parts.**

The use of defined parts for the construction of new pathways is an attractive proposal (Ellis, Adie and Baldwin, 2011). Ribosome binding sites (RBSs) are an essential part for the expression of proteins and can be very useful in modulating gene expression, especially in the construction of synthetic operons (Farasat *et al.*, 2014). However while the core RBS sequence is small, the strength of an RBS can be very context dependent, with the initial ORF affecting the RBS (Espah Borujeni, Channarasappa and Salis, 2014). For this reason having reusable RBS parts for multiple ORFs is challenging (Kaberdin and Bläsi, 2006). The use of a bicistronic design can be used to work around the issue of RBS context dependence (Mutalik *et al.*, 2013), but can cause toxicity issues when multiple bicistronic RBS parts are used.



One approach for creating predictable custom RBSs for each ORF is the use of a biophysical model for translation initiation. This allows RBS strength to be predicted or designed *in silico*, and such a tool is available through a web portal as “the ribosome binding site calculator” (Salis, Mirsky and Voigt, 2009). In practice this tool is not completely reliable, although it maybe allows design in the regions of strong, medium and low strength (Jeschek, Gerngross and Panke, 2016). Another approach has been the use of this tool to facilitate the design of RBS libraries of reduced size which still cover the hypothetical RBS strength space which can be useful in pathway optimization (Jeschek, Gerngross and Panke, 2016).

#### **6.1.4.2 *T. thermophilus* Shine-Dalgarno sequence**

The anti-Shine-Dalgarno (SD) sequence found at the 3' of the *T. thermophilus* 16S rRNA is ACCTCCTTT (Murzina, Vorozheykina and Matvienko, 1988).

Complementarity can be seen with the RBS identified for the S-layer protein, SlpA. This has a sequence of *AAGGAGGTG TGAGGCAT ATG*, with the SD sequence and start codon shown in italics (Faraldo *et al.*, 1992). The anti-SD sequence of *T. thermophilus* is used in the design of custom RBSs using the RBS calculator (Salis, Mirsky and Voigt, 2009).

### **6.1.5 Designing genes for expression in *T. thermophilus***

#### **6.1.5.1 Codon usage**

In most species synonymous codons are not used equally where there is a bias towards certain codons over others. Codons that are not used frequently are classified as rare codons, and are often associated with rare tRNAs where their inclusion in CDS's inhibits protein translation (Gould, Hendy and Papamichail, 2014). *T. thermophilus* has a large number of such rare codons, using 36 % of its codons less than 10 % of the time. Initially in this study the level of codon bias by *T. thermophilus* was not appreciated and native genes were used. No expression was detected for these constructs and codon optimization has proved to be important for protein expression in *T. thermophilus*.

Codon optimisation is the process of trying to match the codon usage of a gene towards the preference of the host, and can be performed in a number of ways

(Gould, Hendy and Papamichail, 2014). Some methods try to maximize the CAI (codon adaption index), although this generally leads to the selection of the most used codon at every position, which can be detrimental to protein expression if certain tRNA pools are exhausted (Gingold and Pilpel, 2011). Another approach is try and emulate a given codon distribution, for example by randomly assigning codons to match the host bias (in many cases with the exception of rare codons which are ignored) (Gould, Hendy and Papamichail, 2014). For the codon optimisation of genes for *T. thermophilus* the random selection of codons to match the host bias, with the removal of rare codons used less than 10 % of the time, was used.

Codon context is also thought to play a role in translation efficiency (Gould, Hendy and Papamichail, 2014). A number of prokaryotes have been shown to have bias towards specific combinations of codons, often referred to as codon pairing. There is evidence that codon pair bias has an influence on translation efficiency (Coleman *et al.*, 2011). Codon reuse has also been observed, with the grouping of specific synonymous codons into distinct regions of some genes. This is thought to allow tRNA recycling and there is evidence that genes that have groups of synonymous codons show better expression (Cannarrozzi *et al.*, 2010). Codon context may be an important consideration for the design of synthetic genes for expression in *T. thermophilus*, however to date this has not been explored.

#### **6.1.5.2 Other factors that may affect translation initiation or elongation in *T. thermophilus*.**

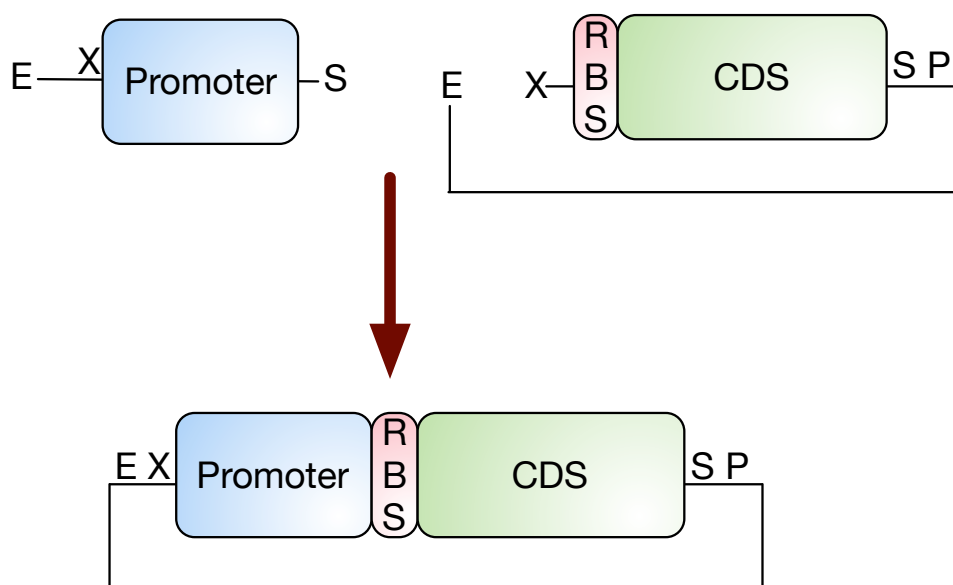
mRNA secondary structure is thought to play a role in translation efficiency. Some secondary structures have been shown to stall ribosomes as they are required to unwind these sections of mRNA before translation elongation can occur (Buchan and Stansfield, 2007). mRNA secondary structure around the area of translation initiation area has also been shown to be detrimental to gene expression (Tuller *et al.*, 2010).

Shine-Dalgarno like sequences inside the CDS have been shown to cause translational pausing in some prokaryotes, due to hybridization between the mRNA and 16S RNA of the ribosome (Li, Oh and Weissman, 2012). Other factors such as repeated nucleotides, polyadenylation sites, nuclease cleavage sites, hidden stop

codons and GC content have also been suggested as factors to consider when designing genes for protein expression (Gould, Hendy and Papamichail, 2014).

### 6.1.6 Assembly of parts using BioBricks cloning approach

The BioBricks cloning approach makes use of four main restriction enzymes to allow modular and iterative cloning. *EcoRI* and *XbaI* are placed upstream of each part, and *SpeI* and *PstI* are placed downstream of each part. *XbaI* and *SpeI* cut sites can be ligated together forming a scar site which cannot be cut by either of these restriction enzymes. This allows the same four restriction sites to be maintained either side of a new combined part for the next round of cloning, as shown in Figure 6-4). Using these restriction sites for all the DNA parts in this study allows their combination and reuse in any combination. Specific BioBricks cloning steps have not been included in the results or methods sections, but all followed standard protocols described here and in the general materials and methods section (section 1.1). All vectors constructed in this study are also available as a supplementary file.



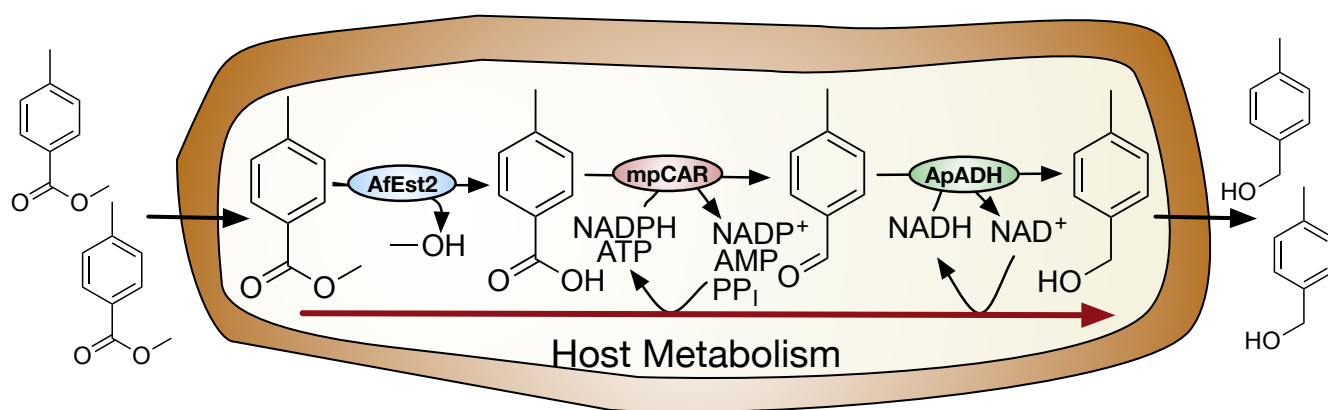
**Figure 6-4 – Standard BioBricks cloning method**

An example of a standard BioBricks cloning procedure. E: *EcoRI*, X: *XbaI*, S: *SpeI*, P: *PstI*. The figure shows a promoter part cut with E and S, and a vector containing a CDS part cut with E and X. These are ligated together to form a new vector containing a newly formed combined part, maintaining the unique restriction enzymes at each side.

### **6.1.7 Summary of aims – a preliminary investigation into the use of *T. thermophilus* as a host organism for whole cell biocatalysis**

This study aims to develop and test the use of *T. thermophilus* as a host organism for whole-cell biocatalysis. The construction of a *T. thermophilus* BioBricks vector will be investigated to facilitate the modular use of DNA parts. Promoters of a range of strengths will be characterised using a reporter protein such as sfGFP. The use of the RBS calculator for the design of custom RBSs for each CDS will be used, and the success of this approach evaluated.

As a proof of principle to investigate the use of *T. thermophilus* for whole-cell biocatalysis, a pathway made up of an esterase from *Archaeoglobus fulgidus* (AfEst2), a carboxylic acid reductase from *Mycobacterium pheli* (mpCAR) and an alcohol dehydrogenase from *Aeropyrum pernix* (ApADH) will to be expressed in *T. thermophilus* (Figure 6-5), utilising the DNA parts described above. Methyl *p*-toluate will be used as a substrate to be hydrolyzed and subsequently reduced through to *p*-tolyl alcohol.



**Figure 6-5 – A *T. thermophilus* cell expressing AfEst2, mpCAR and ApADH, catalyzing the hydrolysis and subsequent reduction of methyl *p*-toluate through to *p*-tolyl alcohol.**

AfEst2: an esterase from *Archaeoglobus fulgidus*, mpCAR: a carboxylic acid reductase from *Mycobacterium pheli*, ApADH: an alcohol dehydrogenase from *Aeropyrum pernix*.

Expression of all three enzymes allows the cascade reaction to be carried out inside the cell with cofactors regenerated from the host metabolism. Substrates and products must diffuse into and out of the cell.

## 6.2 Materials and methods

### 6.2.1 Materials

#### 6.2.1.1 Bacterial strains

Competent 5-alpha *E. coli* (NEB) were used in all DNA manipulations and plasmid preparations. *T. thermophilus* HB27::*nar*, a strain of HB27 with the addition of the *nar* element transferred from *T. thermophilus* HB8, was used in all *T. thermophilus* work. This strain was kindly gifted by José Berenguer, and is required for the use of the Pnar promoter in an HB27 strain.

#### 6.2.1.2 Vectors

Template and cloning vectors used in this study are shown in Table 6-2. Vectors constructed throughout this study for the expression of genes in *T. thermophilus*, are included as a supplementary file.

Table 6-2 – Template and cloning vectors used in this study	
Vector	Description
pMKE2	pMKE2 is a <i>T. thermophilus</i> - <i>E. coli</i> expression vector, purchased from Biotools (Madrid) and used as a template for the construction of pBBTth.
pSB1C3, pSB1A3 and pSB1K3	Established BioBricks plasmids for <i>E. coli</i> available from iGEM ( <i>iGEM website for part BBA_I13521</i> , 2016). They include selectable markers for resistance against chloramphenicol, ampicillin or kanamycin respectively.
pBBTth	The <i>T. thermophilus</i> - <i>E. coli</i> BioBricks vector constructed in this study and used in all <i>T. thermophilus</i> work.
These vectors were either used as a template for the construction of new vectors, or used in the assembly of BioBrick parts. Sequences for other vectors as well as these are available as a supplementary file.	

### 6.2.1.3 DNA BioBrick parts

DNA BioBrick parts used in the study are detailed in Table 6-3. Sequences are also available as a supplementary file.

Table 6-3 – DNA BioBrick parts used throughout this study.		
Name	Source	Description
PslpA	PCR from pMKE2	A constitutive promoter
Prmp	PCR from <i>T. thermophilus</i> genomic DNA	A constitutive promoter
Pnqo	PCR from <i>T. thermophilus</i> genomic DNA	A constitutive promoter
Pnar	PCR from pMKE2	A nitrate and anoxic conditions inducible promoter

AfEst2*	Gene synthesis (IDT)	CDS for expression of AfEst2 protein
ApADH	PCR from a vector containing native ApADH sequence	CDS for expression of ApADH protein. Native sequence
mpCAR	PCR from pNIC28-BSA vector containing mpCAR native sequence	CDS for expression of mpCAR protein. Native sequence.
CtTran-His*	Gene synthesis (IDT)	CDS for expression of CtTran protein. Two copies of this gene were designed, one with a His tag and one with a FLAG tag.
RBS for AfEst2	Primers (Invitrogen)	RBS for AfEst2
RBS for ApADH	Primers (Invitrogen)	RBS for ApADH
RBS for mpCAR	Primers (Invitrogen)	RBS for mpCAR
RBS for CtTran-His	Primers (Invitrogen)	RBS for CtTran-His
ApADH*	Gene synthesis (IDT)	CDS for expression of ApADH. An RBS is included.
ApADH* (N term myc)	Gene synthesis (IDT)	CDS for expression of ApADH with the addition of a N terminal myc tag. An RBS is included.
ATN1-ADH*	Gene synthesis (IDT)	CDS for expression of ATN1-ADH. An RBS is included.
ss- $\gamma$ -lactamase*	Gene synthesis (IDT)	CDS for expression of ss- $\gamma$ -lactamase. An RBS is included.
CtTran-FLAG*	Gene synthesis (IDT)	CDS for expression of CtTran protein. Two copies of this gene were designed, one with a His tag and one with a FLAG tag.
sfGFP*	Gene synthesis (IDT)	CDS for expression of sfGFP. An RBS is included.
mRFP*	Gene synthesis (IDT)	CDS for expression of mRFP. An RBS is included.
mpCAR*	Gene synthesis (MWG)	CDS for expression of mpCAR. An RBS is included.
ancCAR*	Gene synthesis (MWG)	CDS for expression of ancCAR. An RBS is included.
Cas9*	Gene synthesis (Twist)	CDS for expression of Cas9. An RBS is included.
Parts were assembled using BioBricks cloning method to develop and investigate the use of <i>T. thermophilus</i> for whole-cell biocatalysis. A * is used to indicate a codon optimized sequence was used.		



## **6.2.2 Assembly of *T. thermophilus* BioBricks vector and BioBricks parts**

### **6.2.2.1 Vector creation**

Fragments with overhangs for Gibson assembly were designed using SnapGene. Fragments were generated by PCR and where appropriate *DpnI* treated. Fragments were joined in a one hour Gibson assembly reaction at 50 °C, with equimolar concentrations of each fragment used. A 20 µl reaction containing 10 µl of Gibson Assembly Master Mix from NEB was used, following manufacturer's instructions. The assembly was transformed into 5-alpha *E. coli* cells (NEB) on LB agar plates containing 50 µg / ml kanamycin. Red colonies were selected for plasmid purification, and the correct assembly confirmed by restriction digest, PCR and sequencing.

### **6.2.2.2 BioBricks assembly of parts**

BioBrick parts are pieces of DNA with a defined function, such as a promoter, RBS or CDS. BioBrick parts were assembled using standard restriction cloning or three antibiotic assembly (3A) assembly. BioBricks assembly makes use of four restriction sites, two either side of each part. *EcoRI* and *XbaI* are used prefixing a part, while *SpeI* and *PstI* are used suffixing a part. *XbaI* and *SpeI* cut sites can be ligated together forming an uncuttable scar site, maintaining the same four restriction sites either side of a new assembly. The standard prefix sequence is gaattcgcggccgcttctagag, although this is changed to gaattcgcggccgcttctag if the part starts with ATG and is a CDS. The standard suffix sequence is tactagtagcggccgctgcag.

### **6.2.2.3 Creation of RBS parts from primers**

Pairs of complementary forward and reverse primers were ordered from Invitrogen and re-suspended in distilled water at a concentration of 100 µM. Primers were phosphorylated using a T4 polynucleotide kinase (NEB) following the manufacturer's instructions. Primers were mixed to a final concentration of 10 µM each in 30 µl, and 4 µl of 0.5 M NaCl added. Using a thermocycler samples were heated at 95 °C for 2 minutes, followed by 75 cycles of 30 seconds with temperature dropping by 1 °C per

cycle. A 1 in 10 dilution of the primer mix was made, equal to approximately 1 µg of DNA.

#### **6.2.2.4 Assembly of Cas9 by Gibson assembly**

The gene for Cas9 expression in *T. thermophilus* was assembled from three linear pieces of synthetic DNA (IDT), into pSB1C3. Gibson assembly was carried out as described for vector assembly. Cas9 was then cloned into pBBTTh with a range of promoters using standard BioBrick cloning techniques.

#### **6.2.2.5 Codon optimisation for *T. thermophilus***

The codon optimisation tool from IDT was used to match the codon bias of *T. thermophilus* in the creation of synthetic genes. Codons used less than 10 % of the time were ignored, with codons for each amino acid selected randomly weighted by their frequency in the *T. thermophilus* genome.

### **6.2.3 *T. thermophilus* methods**

#### **6.2.3.1 *T. thermophilus* complex media**

*T. thermophilus* was routinely grown on a complex media, often referred to as TM or TB media, as described in (Averhoff, 2006). This was made up of 8 g/L trypticase, 4 g/L yeast extract, 3 g/L NaCl. For TB agar 20 g/L agar was added. Media was titrated to pH 7.5 with HCl and NaOH and autoclaved. A 1 in 1000 dilution of an autoclaved 1000x MgCl and CaCl<sub>2</sub> solution containing 7.2 g/L MgCl<sub>6</sub>H<sub>2</sub>O and 8.7 g/L CaCl<sub>2</sub>·2H<sub>2</sub>O was added, as these ions have been shown to be important in transformation efficiency (Averhoff, 2006).

#### **6.2.3.2 *T. thermophilus* minimal media**

For promoter characterization a minimal media was required to reduce background fluorescence. This was made up of 20 g/L sucrose, 20 g/L sodium glutamate, 0.5 g/L K<sub>2</sub>HPO<sub>4</sub>, 0.5 g/L KH<sub>2</sub>PO<sub>4</sub>, 0.25 g/L NaCl, 0.5 g/L (NH<sub>4</sub>)SO<sub>4</sub> with pH titrated to between 7.0 - 7.2. After autoclaving a 1 in 1000 dilution of the 1000x MgCl and CaCl<sub>2</sub> solution described previously was added. A 1 in 1000 dilution of a filter sterilised solution, made up in 0.01 N H<sub>2</sub>SO<sub>4</sub> containing, 60 g/L FeSO<sub>4</sub>·7H<sub>2</sub>O, 8 g/L CoCl<sub>2</sub>·6H<sub>2</sub>O, 0.2 g/L NiCl<sub>2</sub>·6H<sub>2</sub>O, was also added. The addition of 0.1 mg/L biotin

and 1 mg/L thiamine is also suggested in some literature (Averhoff, 2006), however these were not added in this study.

#### **6.2.3.3 *T. thermophilus* growth**

As standard *T. thermophilus* was grown in 20 to 30 ml of TB media, in a 250 ml conical flask, at 70 °C with shaking at 200 RPM. Where *T. thermophilus* was transformed with a plasmid containing a kanamycin resistance cassette, 30 µg/ml kanamycin was added. Glycerol stocks were produced as they were for *E. coli* cultures, through the addition of 20 % glycerol to a culture before freezing at -80 °C. Confluent cultures could be obtained overnight through inoculation 30 ml TB media with a *T. thermophilus* colony or small scrape of a glycerol stock. Complete coverage of the conical flask lid with tin foil over a bung was important to prevent high levels of evaporation.

#### **6.2.3.4 *T. thermophilus* transformation**

*T. thermophilus* was grown overnight in TB media at 70°C, 200 RPM before being diluted to OD<sub>600nm</sub> 0.1. 5 µl of plasmid DNA was added to 500 µl of *T. thermophilus* culture in 10 ml breather tubes and incubated for four hours at 70 °C, 200 RPM. Cultures were plated on TB agar with 30 µg/ml kanamycin. Plates were incubated at 70 °C inside a sealed plastic box to which a small volume of water was added in a Duran bottle lid, which is important to prevent the agar plates from becoming dehydrated.

#### **6.2.3.5 Promoter characterisation**

Overnight cultures of *T. thermophilus* transformed with plasmids for the expression of sfGFP from the range of promoters were set up in TB media as standard. Minimal media cultures were inoculated with 500 µl of the overnight culture in 30 ml minimal media with 30 µg/ml kanamycin. Cultures were incubated at 60 °C, 200 RPM. 200 µl samples were taken in triplicate into a 96-well microtiter plate for OD<sub>600</sub> and fluorescence readings. Fluorescence settings were as follows: Excitation wavelength 450 nm, excitation bandwidth 9 nm, emission wavelength 508 nm, emission bandwidth 20 nm, gain 120, 10 flashes, 10 µs integration time, lag time 0 µs, settle time 0 µs, and an automatically calculated Z-position of 18268 µm. Population average GFP fluorescence was calculated. This is blank subtracted

fluorescence divided by blank subtracted OD<sub>600nm</sub>. Standard deviations for each reading were combined and shown as error bars.

#### **6.2.3.6 Whole-cell biocatalysis assays**

*T. thermophilus* was grown as specified previously. 250 mM stocks of each substrate was prepared in ethanol and added to give the desired concentration (usually 2 mM). Conversion of substrates by *T. thermophilus* was determined by HPLC. 500 µl samples were taken and added to 500 µl of acetonitrile. Samples were centrifuged at maximum speed for 10 minutes, and the supernatant collected. Samples were kept at 4 °C and analyzed by HPLC within 24 hours. The standard HPLC method specified in the general materials and methods chapter was used.

## 6.3 Results

### 6.3.1 Construction of a *T. thermophilus* BioBricks vector.

#### 6.3.1.1 Specifications for a *T. thermophilus* BioBricks vector

In order to use the BioBricks cloning approach for the assembly of synthetic biology parts for expression in *T. thermophilus*, a BioBricks vector was needed. A vector capable of replication and selection in both *E. coli* and *T. thermophilus*, with an insulated BioBricks cloning site, was required. Minimization of vector size was also considered an important goal.

#### 6.3.1.2 Template vectors

Two vectors were selected as templates for the construction of a new vector (Figure 6-6). pMKE2 is a commercially available vector derived from pMK18 (Figure 6-6A) which meets many of the criteria above. However it contains a number of regions of apparently “junk” DNA and contains a limited cloning site only for the expression of genes from its Pnar2 promoter. pSB1C3 (Figure 6-6B) is an established BioBricks vector for *E. coli* with a BioBricks cloning site flanked by two insulating terminator regions. The BioBricks cloning site is supplied containing a stuffer gene for the expression of mRFP under the control of a lac promoter which helps identify successful cloning attempts.



### **Figure 6-6 – Template vectors used for the construction of both *T. thermophilus* BioBricks vectors**

These vectors were used for the construction of new *T. thermophilus* vectors by PCR and Gibson assembly. Fragments amplified for the construction of *T. thermophilus* BioBricks vector 1 are shown by the red lines. Fragments amplified for the construction of *T. thermophilus* BioBricks vector 2, called pBBTTh, are shown by green lines.

**A:** pMKE2 - An *E. coli* – *T. thermophilus* shuttle vector with a Pnar promoter for the expression of proteins in *T. thermophilus*. Important features include:

PslpA-KanR - A heat stable kanamycin resistance cassette, under the control of a bi-functional PslpA promoter.

RepA - A region for replication in *T. thermophilus*.

Ori – Region of replication for *E. coli*.

pNAR2 – A modified pNAR promoter for driving protein expression in *T. thermophilus*

**B:** pSB1C3 – A BioBricks vector for *E. coli*. Important features include:

CmR – A chloramphenicol resistance cassette

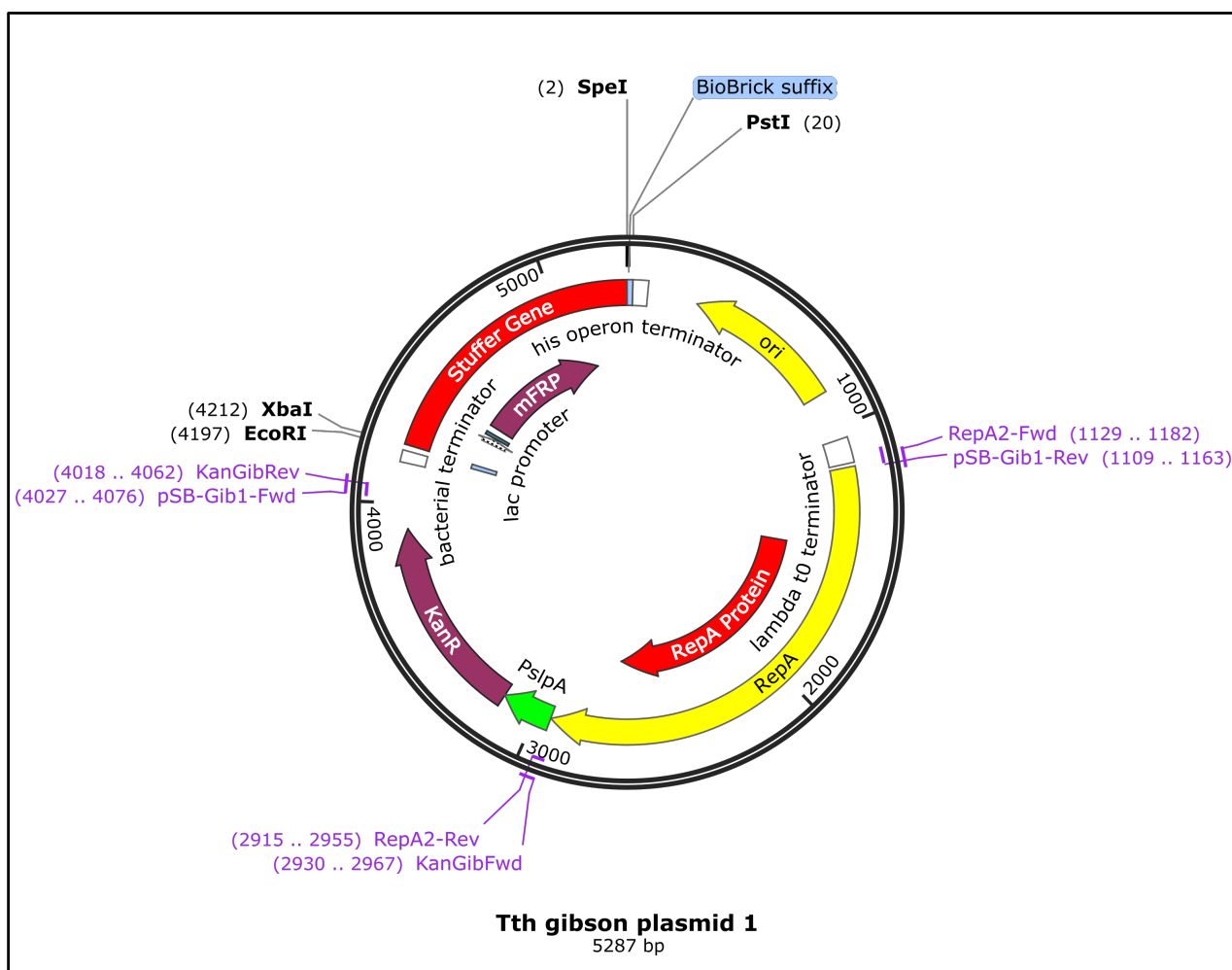
BioBricks cloning site: Restriction sites *EcoRI*, *XbaI*, *SpeI* and *PstI* flanking stuffer gene for the expression of mRFP. Terminators flank this cloning site.

Ori – Region of replication for *E. coli*. Identical to the same region in pMKE2

#### **6.3.1.3 *T. thermophilus* BioBricks vector 1 construction**

The first vector constructed (Figure 6-7) took the BioBricks cloning site and origin of replication for *E. coli* from pSB1C3 as one fragment (Figure 6-6B – Red line), and the RepA region for replication in *T. thermophilus* and bi-functional kanamycin resistance cassette as two separate fragments from pMKE2 (Figure 6-6A – Red lines). Fragments were amplified by PCR with overhangs to allow Gibson assembly.

This first vector had a minimized size and was able to function in *E. coli*. Red colonies were selected from agar plates containing 50 µg / ml kanamycin, and the new vector confirmed by PCR and restriction digest (data not shown). However this vector was unable to be transformed into *T. thermophilus* (data not shown). The RepA region of this vector was also confirmed by sequencing.



**Figure 6-7 – *T. thermophilus* BioBricks plasmid 1**

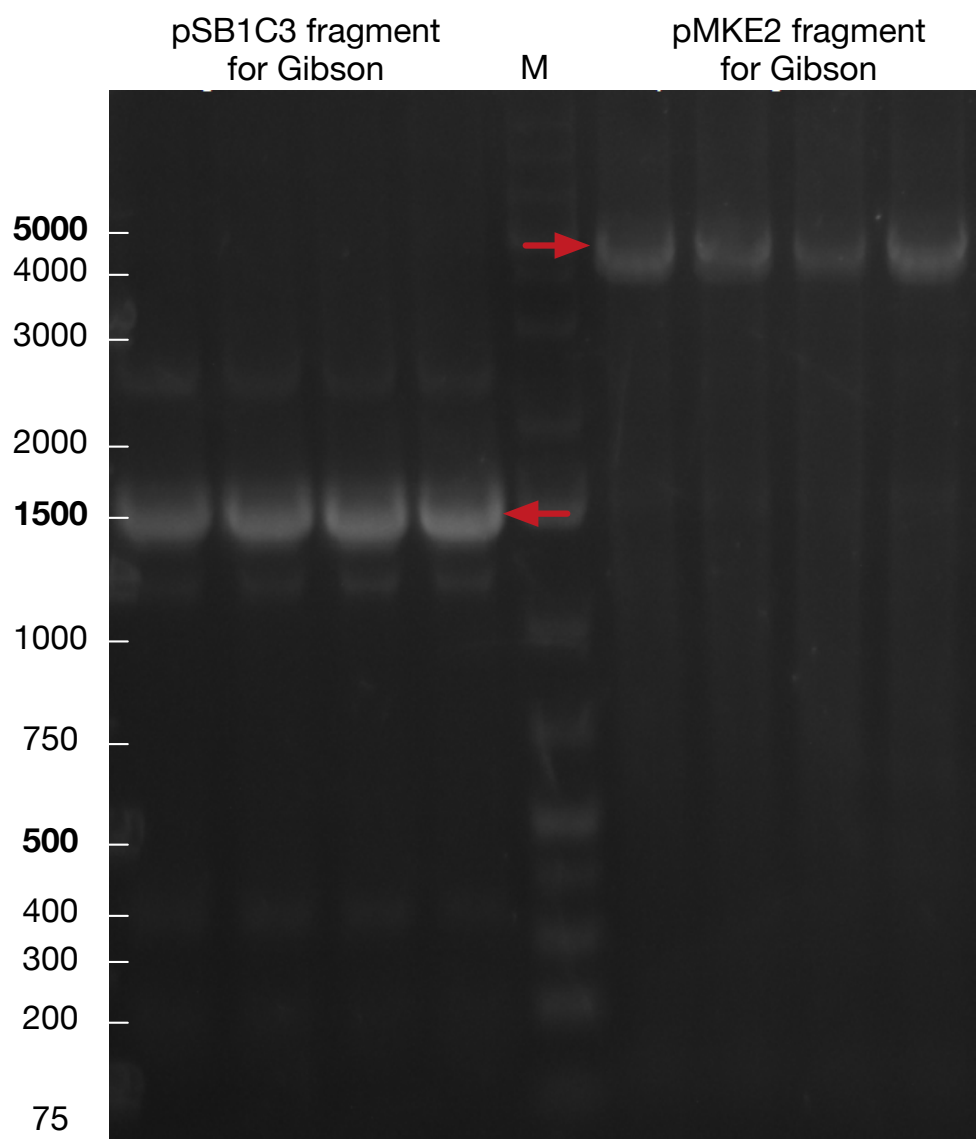
The first attempt at constructing a *T. thermophilus* BioBricks vector. This vector has regions for replication in both *E. coli* and *T. thermophilus*, and a bi-functional and heat stable kanamycin resistance cassette. Vector size was minimized by not including sections of DNA from the template vectors with no apparent purpose. This vector did not function in *T. thermophilus*.



#### **6.3.1.4 *T. thermophilus* BioBricks vector 2 construction - pBBTTh**

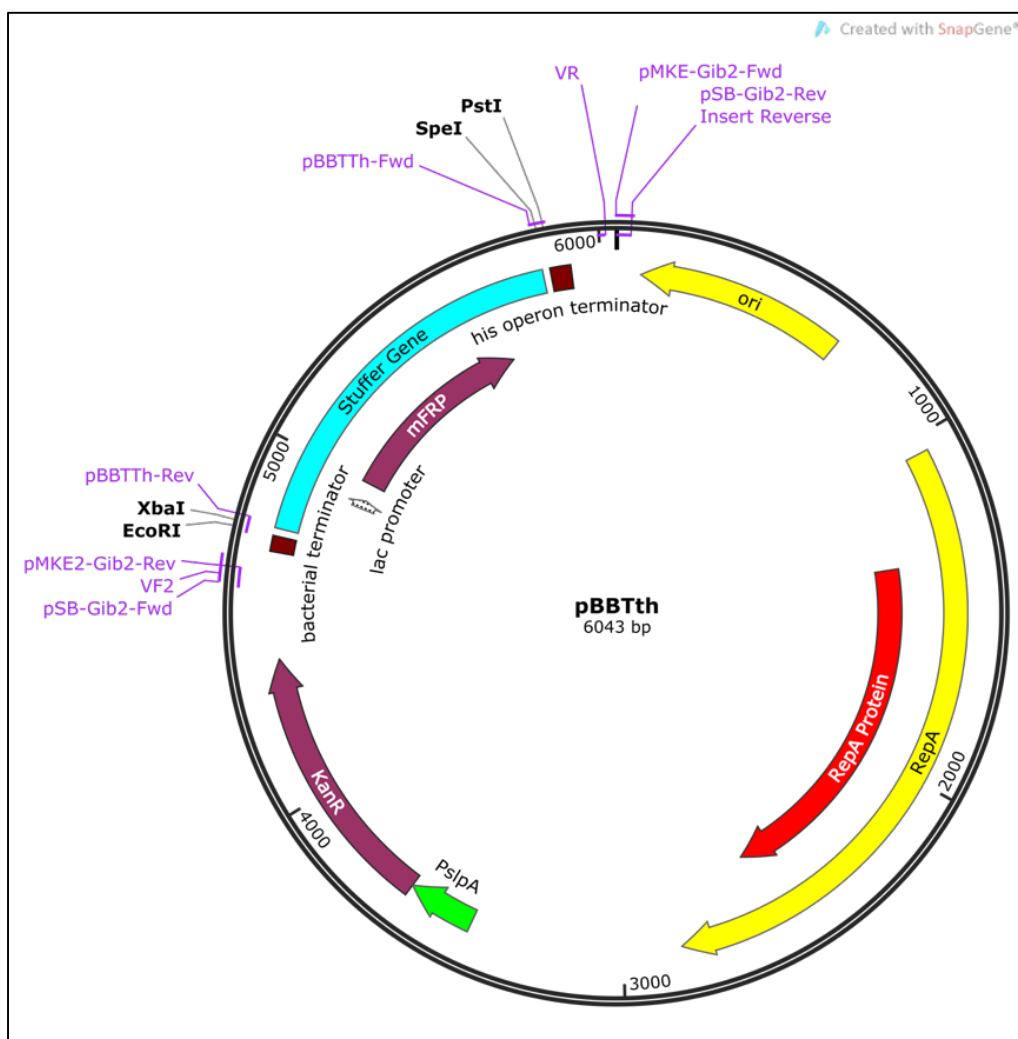
A second vector was designed (Figure 6-9), without the removal of so much of the pMKE2 vector DNA, as this vector is known to function in *T. thermophilus*. A fragment containing the BioBricks cloning site was taken from pSB1C3 (Figure 6-6B – Green line), and a fragment containing everything else from pMKE2 (Figure 6-6A – Green line). Again these fragments were amplified by PCR with overhangs to allow Gibson assembly (Figure 6-8).

Again red *E. coli* colonies were selected on 50 µg / ml kanamycin (Figure 6-10 – Part I). This vector was also able to function in *T. thermophilus*, with transformants able to grow on 30 µg / ml kanamycin (Figure 6-10 – Part II). The correct assembly was confirmed by restriction digest using *Xba*I and *Pst*I, with bands at the correct size identified (Figure 6-11). This vector was designated pBBTTh and used in subsequent work (Figure 6-9).



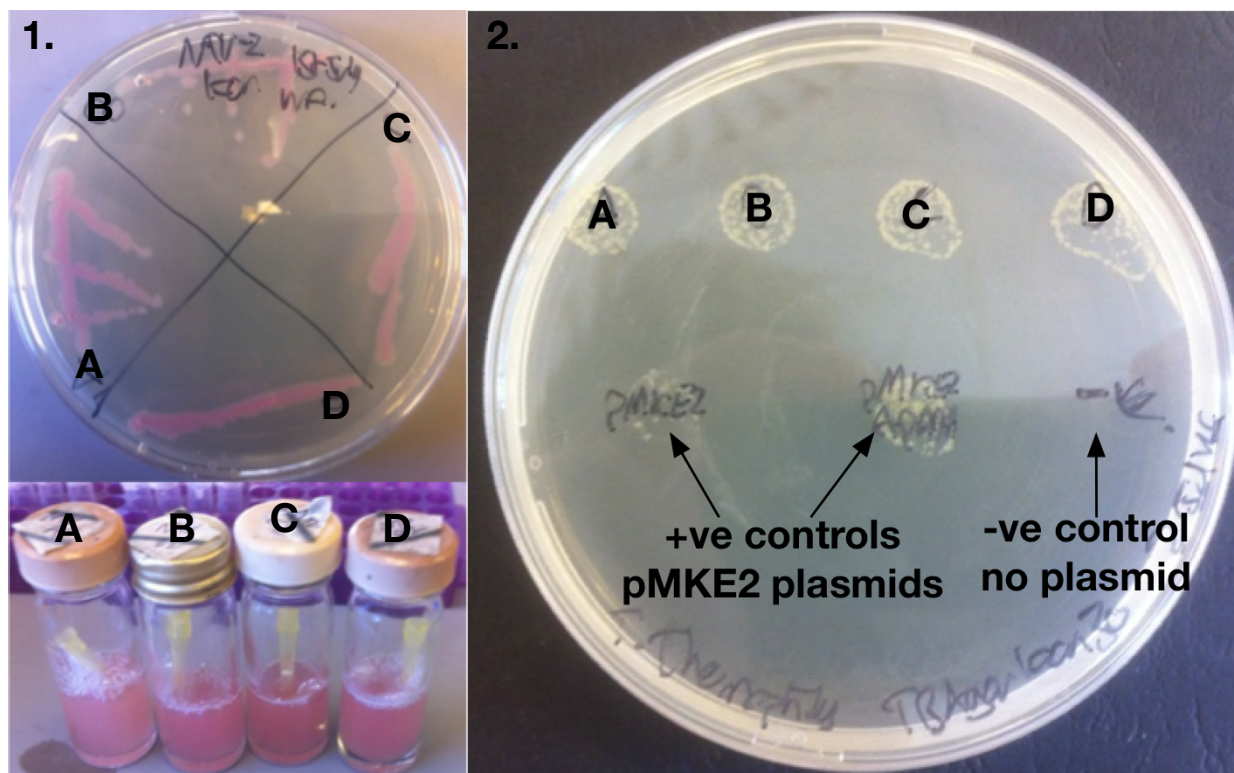
**Figure 6-8 – PCR amplification of two fragments for the construction of a second *T. thermophilus* BioBricks vector by Gibson assembly.**

Fragments were amplified from pMKE2 and pSB1C3 (Figure 6-6 – Green lines) using a temperature gradient from 55 °C to 75 °C (left to right on for both fragments on the gel). However all temperatures resulted in equally selective PCR, as shown on the gel. Bands are at the expected size for the fragments amplified, indicted by arrows in the figure.



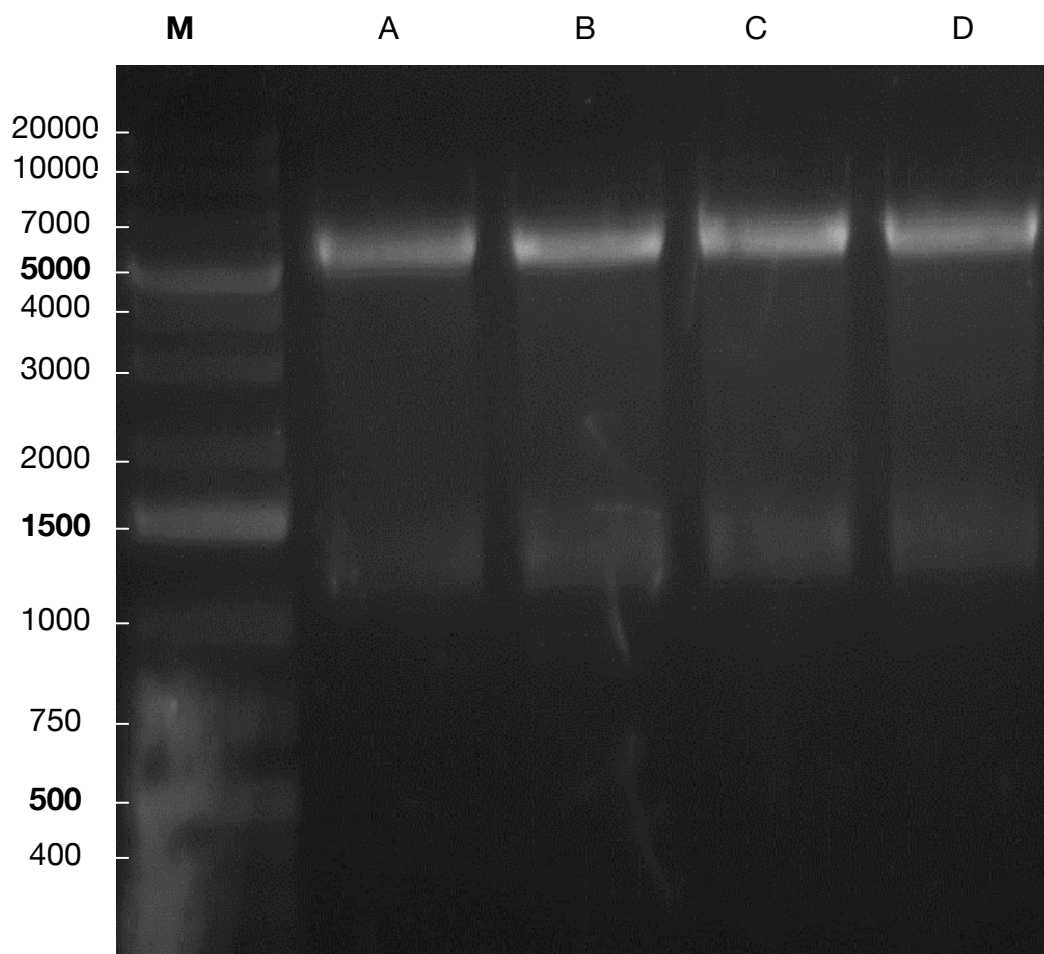
**Figure 6-9 – *T. thermophilus* BioBricks plasmid 2 - pBBTth**

The second attempt at constructing a *T. thermophilus* BioBricks vector. This vector was constructed by Gibson assembly by combining the PCR fragments in Figure 6-8. Again, this vector has regions for replication in both *E. coli* and *T. thermophilus*, and a bifunctional and heat stable kanamycin resistance cassette. However sections of DNA from the pMKE2 with no apparent purpose were not removed, and this vector was functional in *T. thermophilus*, unlike the first vector.



**Figure 6-10 – pBBTTh functioning in both *E. coli* and *T. thermophilus*.**

1. Four *E. coli* colonies were picked labelled A to D. Each was red due to the expression of mRFP in the stuffer gene, and able to grow in the presence of 50  $\mu\text{g}$  / ml kanamycin from the bifunctional kanamycin resistance cassette.
2. Plasmids were isolated from *E. coli* colonies A to D and transformed into *T. thermophilus*. Transformants were able to grow in the presence of 30  $\mu\text{g}$  / ml kanamycin in all cases showing the vector was functional. No growth was detected in a negative control of *T. thermophilus* without the addition of a plasmid.



**Figure 6-11 – Restriction digest of pBBTTh by *XbaI* and *PstI***

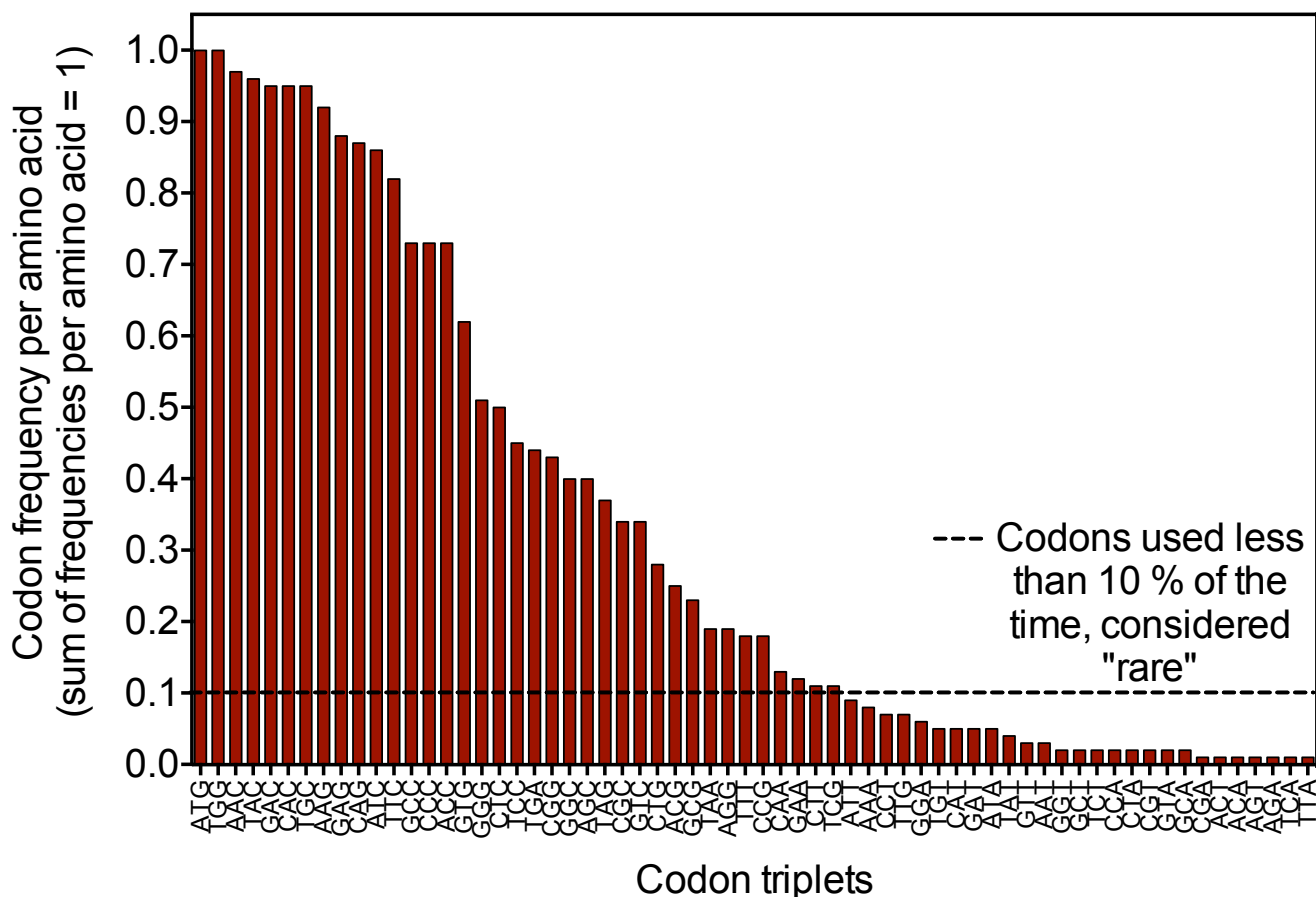
A restriction digest of pBBTTh was carried out using restriction enzymes *XbaI* and *PstI*. Bands of the correct sizes were identified on the gel confirming the successful construction of the second *T. thermophilus* BioBricks vector. A-D shows four separate plasmid samples identical to those used in Figure 6-10, all digested with *XbaI* and *PstI*.

### **6.3.2 Synthetic biology tools for controllable protein expression in *T. thermophilus*.**

#### **6.3.2.1 Codon usage in *T. thermophilus* is limited due to high GC content.**

*T. thermophilus* has a GC content of 70 %, and as a result its codon usage is more limited than average. The relative frequency a codon is used to code for a specific amino acid can be calculated as a fraction of use, summing to one for each amino acid (Figure 6-12). Rare codons can arbitrarily be defined by specifying a

percentage usage cut-off, with 10 % being used by some gene synthesis companies, such as IDT, as rare codons to avoid in codon optimization. Taking this cut-off for *T. thermophilus* codon usage, 36 % of the codons are classed as rare. The strict codon usage of *T. thermophilus* makes codon optimization almost essential for successful heterologous gene expression. In the initial design of some constructs this was not appreciated and the native genes used (Figure 6-16). No expression was detected in these cases (Figure 6-19 and Figure 6-21)



**Figure 6-12 - *Thermus thermophilus* HB27 codon usage**

Codon usage was taken from the Kazusa database (Nakamura, Gojobori and Ikemura, 1999). Codon frequency per amino acid is shown, which equals a total of 1 for each amino acid. A threshold of 0.1 is shown which equates to triplets which are used less than 10 % of the time. Codons used less than this may be considered as “rare” and equate to 36 % of the possible triplets.

#### **6.3.2.2 Ribosome binding site calculator for designing maximised RBS's for *T. thermophilus*.**

Custom ribosome binding sites (RBS) were designed for each coding sequence (CDS), using the RBS calculator. A maximization of RBS strength was set as the target objective. Initially RBS's were designed as BioBricks themselves, with the aim of being able to easily modulate RBS strength by swapping out the RBS part. However this proved to be an inefficient use of time, and moreover the resulting scar site between the RBS and CDS limited the ability of the RBS calculator in maximizing RBS strength. RBS's that were designed as BioBrick parts, and those that were designed with no scar site constraints, are indicated in Table 6-4. In general the use of the RBS calculator for designing maximized RBS's for *T. thermophilus* was quite successful.

<b>Table 6-4 – Ribosome binding site strength</b>	
Gene	RBS strength (A.U.)
AfEst2* <sup>C</sup>	69,963
ApADH <sup>C</sup>	170,863
ApADH* <sup>C</sup>	19,141
ApADH* (N term myc)	626,778
ATN1-ADH*	50,417
ss- $\gamma$ -lactamase*	42,111
mpCAR	40,768
CtTran-His* <sup>C</sup>	50,417
sfGFP*	243,592
mRFP*	1,122,434
CtTran-FLAG*	203,462
mpCAR*	135,374
ancCAR*	1,287,767
Cas9*	305,063
<p>RBS strengths for all of the proteins expressed in <i>T. thermophilus</i> in this study, as calculated using the RBS (Salis, Mirsky and Voigt, 2009). (A.U. - Arbitrary Units). The scale runs from 0 to 5,000,000. * represents a codon optimized sequence. Initial designs used individual BioBrick RBS parts, resulting in constraints on the RBS design, this is indicated by <sup>C</sup>.</p>	



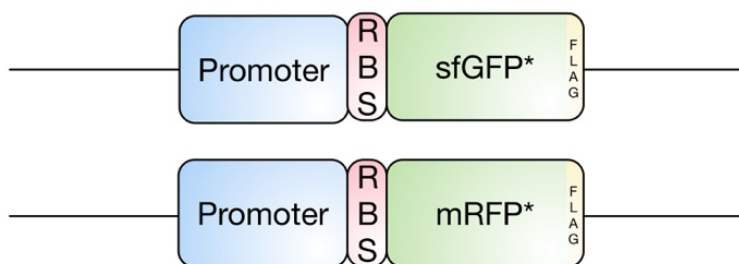
### 6.3.2.3 Promoters for *T. thermophilus*

Four promoters for *T. thermophilus* were identified as covering a range of strengths, partly through personal communication with Prof. José Berenguer and Prof. Aurelio Hidalgo. PslpA and Pnar were cloned as BioBricks from pMKE2 by PCR, while Prmp and Pnqo were cloned from *T. thermophilus* genomic DNA. In an effort to ensure promoters from the *T. thermophilus* genome were captured, 200 bp and 150 bp regions upstream of the *rmpB* and *nqo* genes were used, although it is possible the actual promoter regions are smaller.

### 6.3.2.4 Testing promoter strength using sfGFP

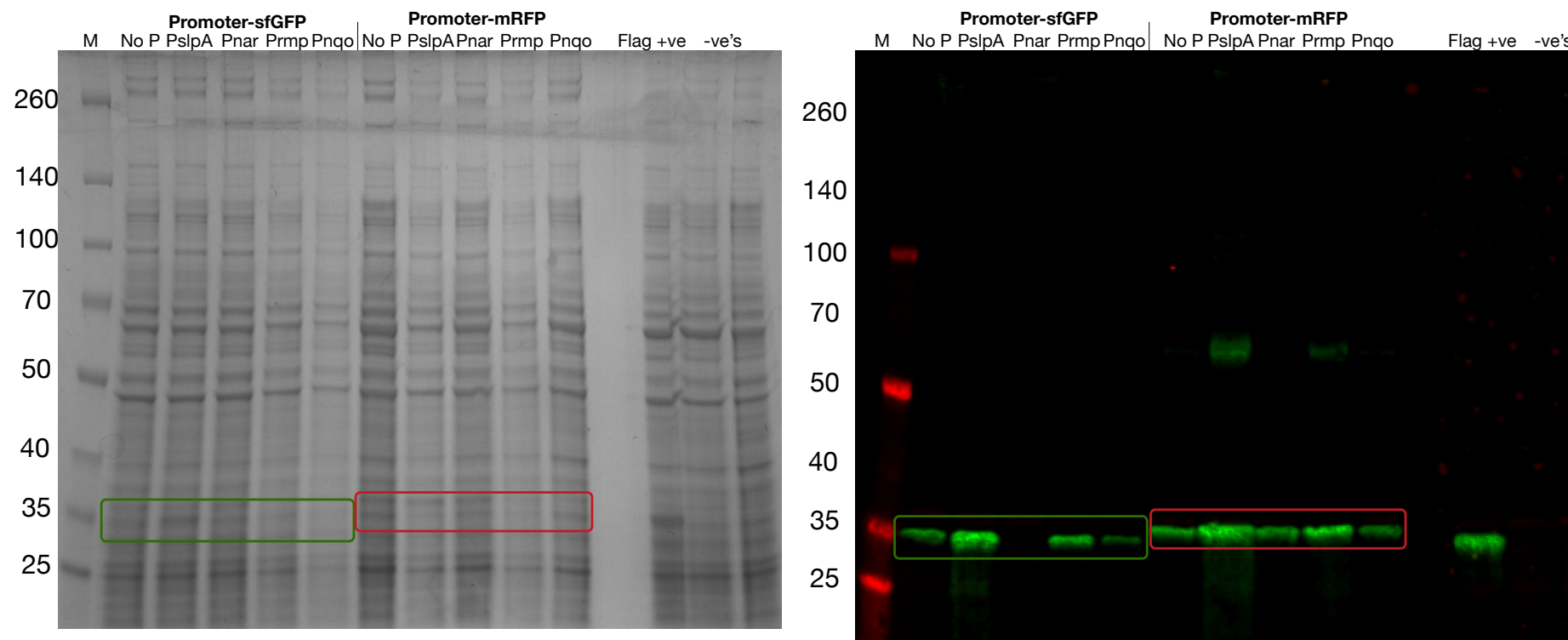
In order to quantify promoter strength, a super-folder GFP (sfGFP) was used able to function at 70 °C (Cava *et al.*, 2008). Evidence that mRFP might function at elevated temperatures was also found through the iGEM website (*iGEM website for part BBa\_I13521*, 2016), although efforts to use mRFP to quantify protein expression in *T. thermophilus* were not effective (data not shown). *T. thermophilus* was transformed with constructs in pBBTth containing each promoter upstream of sfGFP or mRFP (Figure 6-13), as well as with no promoter, and grown in minimal media. Both sfGFP and mRFP were codon optimized for expression in *T. thermophilus* and featured a C-terminal FLAG tag.

To confirm sfGFP and mRFP expression, an SDS-PAGE gel and western blot were run of total cell lysates of these *T. thermophilus* constructs. FLAG tagged protein of the correct molecular weight was clearly detected by western blot analysis, and was faintly detectable by SDS-PAGE (Figure 6-14). To quantify sfGFP expression, cultures grown in minimal media were set up and incubated at 60 °C, 200 RPM. 200 µl samples were taken in triplicate into a 96-well microtiter plate to allow reading of fluorescence and OD<sub>600nm</sub>. At approximately OD<sub>600nm</sub> 0.15 population average sfGFP expression was determined, quantifying gene expression from each promoter (Figure 6-15). Population average sfGFP expression after 24 hours incubation, with an OD<sub>600nm</sub> 1.1 showed similar results (data not shown). The Pnar promoter was not induced and so showed no expression. A blank reading of *T. thermophilus* containing empty pBBTth was subtracted from the population average fluorescence of all other readings.



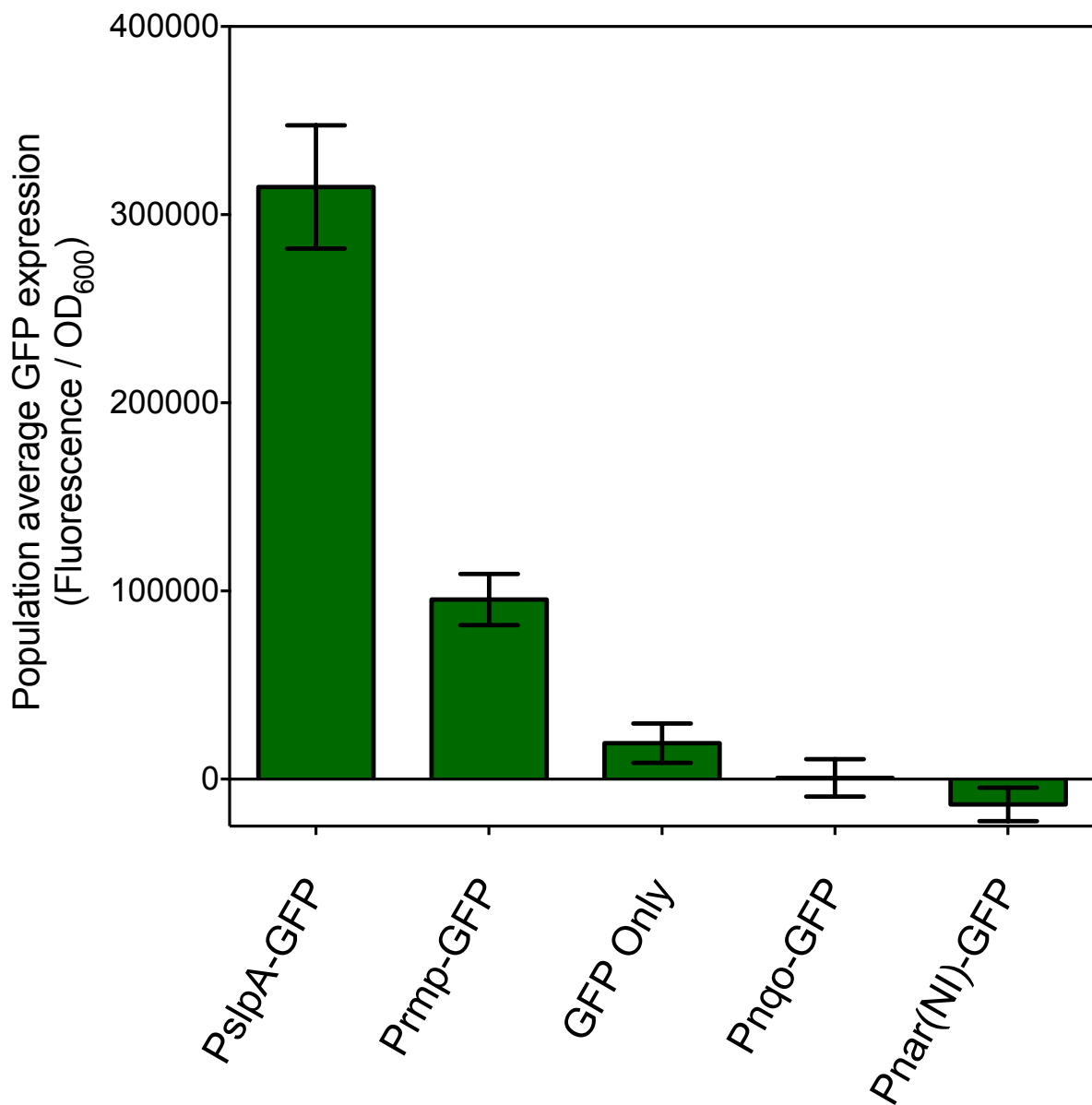
**Figure 6-13 – The constructs used for measuring promoter strength**

From RBS-CDS parts, constructs were made for the expression of sfGFP or mRFP from a range of promoters, as well as without any promoter at all. These constructs were cloned into pBBTTh to allowing testing of promoter strength by measuring sfGFP or mRFP expression by fluorescence. Both CDS's featured a C terminal FLAG tag to allow detection of expression by western blot. \* is used to indicate a codon optimized sequence. Promoter: A promoter part, either PslpA, Pnar, Prmp, Pnqo, or no promoter at all. RBS: A custom RBS, designed using the RBS calculator.



**Figure 6-14 – SDS-PAGE analysis and western blot of total cell lysate of *T. thermophilus* expressing sfGFP and mRFP under the control of a range of promoters.**

SDS-PAGE and western blot of total cell lysate from *T. thermophilus* expressing codon optimised sfGFP and mRFP, under the control of PslpA, Pnar, Prmp, Pnqo and without a promoter. *T. thermophilus* containing empty pBBTthK and without a plasmid were included as negative controls. A *T. thermophilus* cell lysate expressing FLAG tagged AfEst2 was included as a positive control. The western blot was probed using a chicken anti-FLAG primary antibody and a goat anti-chicken IRDye 800CW secondary antibody. The expected size of sfGFP and mRFP is shown by the green and red boxes respectively.



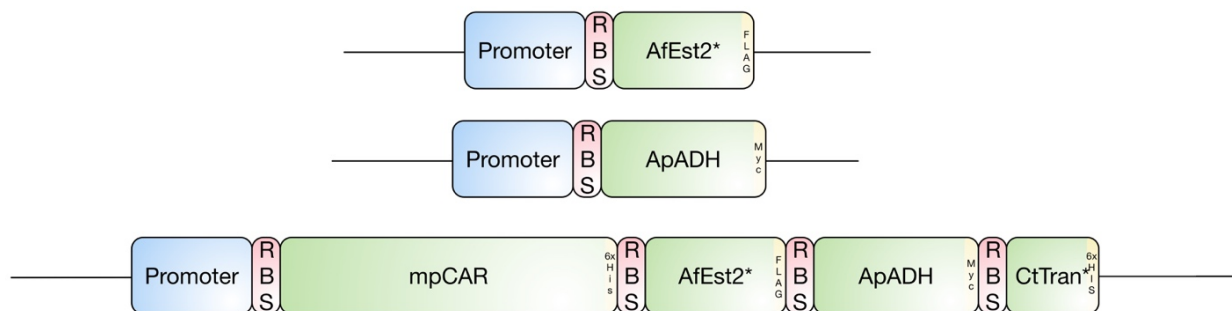
**Figure 6-15 – Population average sfGFP fluorescence as a measure of promoter strength**

sfGFP expression was measured by reading the fluorescence and OD<sub>600nm</sub> *T. thermophilus* were cultures grown in minimal media and 200 µl was taken in triplicate from a shake flask into a 96-well microtiter plate for reading using a plate reader. Population average GFP expression was calculated as the Fluorescence recorder divided by OD<sub>600nm</sub>. A blank was also measured using *T. thermophilus* cultures with no sfGFP expression, and subtracted from all readings. Error bars shows the combined standard deviation for all readings.

### **6.3.3 Initial trials for the expression of AfEst2, ApADH and a synthetic operon.**

#### **6.3.3.1 Building constructs for the expression AfEst2, ApADH and a synthetic operon in *T. thermophilus*.**

Constructs for expression of AfEst2, ApADH and for the expression of a complete synthetic operon (Figure 6-16), were constructed using the BioBricks cloning approach into pBBTth (Figure 6-9). In the synthetic operon a gene for the expression of a phosphopantetheine (PPT) transferase from the thermophile *Clostridium thermocellum* was included (CtTran), for transfer of the PPT group onto a CAR. AfEst2 and CtTran were codon optimised and synthesised by Integrated DNA Technologies (IDT) to match the *T. thermophilus* codon usage, as the native sequences contained BioBricks restriction sites. Native sequences were used for ApADH and mpCAR, although it was subsequently realized that codon optimization is almost essential for successful gene expression in *T. thermophilus* (Figure 6-12). The PslpA, Pnar, Prmp and Pnqo promoters were used for all constructs.



**Figure 6-16 – Constructs for the expression of AfEst2, ApADH and a complete synthetic operon.**

From RBS-CDS parts constructs were made for the expression of AfEst2 or ApADH from a range of promoters. A synthetic operon was also constructed for the expression of mpCAR, AfEst2, ApADH and CtTran from a single promoter. \* is used to indicate a codon optimized sequence.

Promoter: A promoter part, either PslpA, Pnar, Prmp or Pnqo.

RBS: A custom RBS, designed using the RBS calculator.

AfEst2\*: Codon optimized gene for the expression of AfEst2, a C terminal FLAG tag was included.

ApADH: Natural gene for the expression of ApADH, a C terminal Myc tag was included.

mpCAR: Natural gene for the expression of mpCAR, a C terminal 6xHis tag was included

CtTran\*: Codon optimized gene for the expression of CtTran, a C terminal 6xHis tag was included.

### 6.3.3.2 Testing constructs for the expression AfEst2, ApADH.

*T. thermophilus* transformed with constructs for the expression of AfEst2 or ApADH in pBBTth were grown overnight in TB media + 30 µg / ml kanamycin at 70 °C at 200 RPM. Negative controls of *T. thermophilus* with an empty pBBTth plasmid, and without a plasmid were also set up (without kanamycin added where no plasmid was used). Fresh cultures were then inoculated at OD<sub>600nm</sub> 0.05 and growth followed using OD<sub>600nm</sub> (Figure 6-17 and Figure 6-18).

After approximately 3.5 hours one of the cultures containing constructs featuring the Pnar promoter was placed in a sealed and para-filmed 50 ml falcon tube with the

addition of 40 mM potassium nitrate. This was incubated at 70 °C with no shaking to bring about anaerobic conditions. According to previous work this should cause induction of the Pnar promoter (Hidalgo *et al.*, 2004). In both cases a further culture featuring the Pnar promoter was kept shaking at 200 RPM. At 6.5 hours 2 mM methyl *p*-toluate to test Af-Est2, or 2 mM *p*-tolualdehyde to test ApADH, was added as indicated (Figure 6-17 and Figure 6-18).

After 21 hours 500 µl samples were taken for HPLC analysis, and added to 500 µl of acetonitrile. The cell debris was removed by centrifugation, and the remaining supernatant quantified by HPLC against standard curves (Figure 6-17 and Figure 6-18). Samples were also taken at this time for SDS-PAGE and western blot analysis (Figure 6-19).

#### **6.3.3.3 AfEst2 expression in *T. thermophilus*, and the conversion of methyl *p*-toluate.**

AfEst2 was expressed well by PslpA, Prmp and Pnqo, visible by both SDS-PAGE and western blot analysis (green bands). A low level of expression by the induced Pnar promoter was also detected by western blot analysis (Figure 6-19). The western blot was probed using a chicken anti-FLAG and mouse anti-myc primary antibodies. Two secondary antibodies were used, rabbit anti-mouse IRDye 680 and goat anti-chicken IRDye 800CW. *T. thermophilus* containing empty pBBTthK and without a plasmid were included as negative controls, and non-specific binding of multiple antibodies to a protein at approximately 100 kDa was detected in all *T. thermophilus* samples.

Expression levels of AfEst2 seemed to match the promoter strengths determined using sfGFP expression (Figure 6-15), with PslpA showing the most expression followed by Prmp and Pnqo.

While the negative controls showed some conversion of methyl *p*-toluate, *T. thermophilus* expressing AfEst2 showed a much higher conversion to *p*-toluic acid (Figure 6-17). Expression from Prmp gave the best conversion, followed by Pnqo and then PslpA. Conversion of methyl *p*-toluate seems not only to be related to protein expression but also to cell density. All of the initial 2 mM methyl *p*-toluate was not accounted for and it seems likely that a side reaction is occurring.

#### **6.3.3.4 ApADH expression in *T. thermophilus*, and the conversion of *p*-tolualdehyde**

No expression of ApADH was detected by either SDS-PAGE or western blot analysis (Figure 6-19), which is likely due to the strict codon usage of *T. thermophilus* (Figure 6-12). The western blot was probed as described for AfEst2 (section 6.3.3.3). The *p*-tolualdehyde was converted predominantly into *p*-toluic acid, with some *p*-tolyl alcohol, by all the *T. thermophilus* cultures (Figure 6-18). It was hoped that expression of ApADH could allow better conversion into *p*-tolyl alcohol rather than predominantly *p*-toluic acid. The next step was therefore codon optimization of ApADH to match the *T. thermophilus* codon usage.

#### **6.3.3.5 Synthetic operon expression in *T. thermophilus*, and the conversion of methyl *p*-toluate.**

Cultures to test the expression of the synthetic operon (Figure 6-16) were set up in the same way as those to test AfEst2 and ApADH expression. However induction of Pnar was carried out instead at 9 hours due to slower growth, at which point 2 mM methyl *p*-toluate was also added to all cultures. Samples for HPLC, SDS-PAGE and western blot were taken as before, after 23 hours.

Only expression of AfEst2 was detected by SDS-PAGE and western blot analysis (Figure 6-21). The western blot was probed using chicken anti-flag, mouse anti-myc and mouse anti-6x-His primary antibodies. Two secondary antibodies were used, rabbit anti-mouse IRDye 680 (shown in red) and goat anti-chicken IRDye 800CW (shown in green). A band at approximately 45 kDa was detected by the rabbit anti-mouse IRDye 680 antibody in the PslpA sample, however this band is also faintly visible in the negative control suggesting it is not apADH. Myc and 6x-his positive controls were included, although the 6x-his control was not detected.

It is worth noting that the mpCAR gene is approximately 3.5 kbp in length, but that there is good translation of AfEst following this from the internal RBS. AfEst2 expression was to a similar level to AfEst2 expressed alone (Figure 6-19 and Figure 6-21). The mpCAR, ApADH or CtTran expression could not be detected by SDS-PAGE or western blot. However the positive 6x-His positive control was also poorly detected. *T. thermophilus* containing empty pBBTth and without a plasmid were

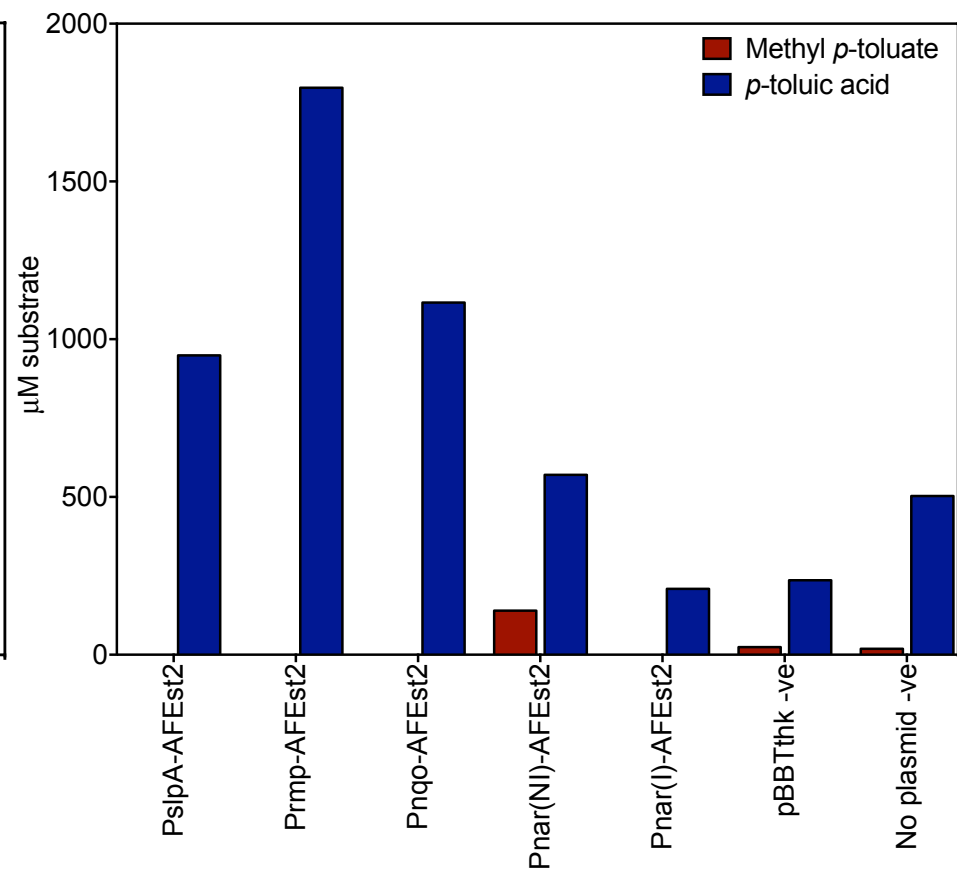
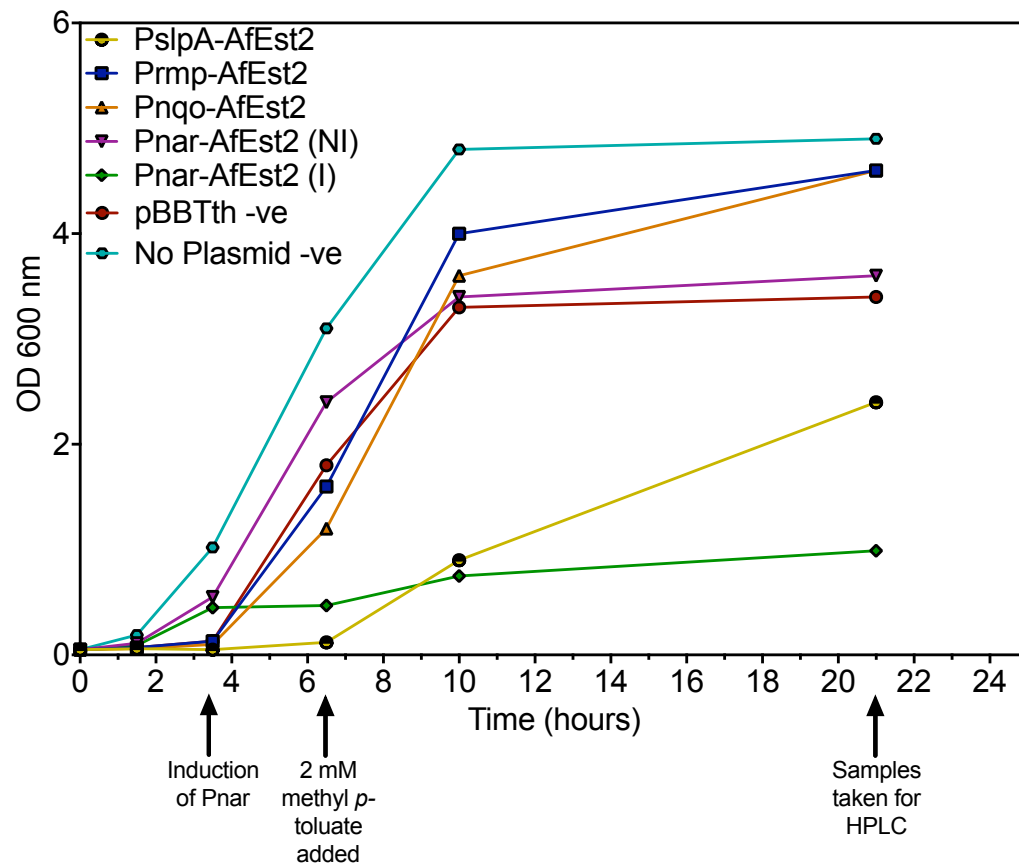


included as negative controls, and non-specific binding of multiple antibodies to a protein at approximately 100 kDa was detected in all *T. thermophilus* samples.

A significant amount of methyl *p*-toluate was converted to *p*-toluic acid by *T. thermophilus* expressing AfEst2, although low levels of *p*-toluic acid were also detected in the empty pBBTth and no plasmid controls (Figure 6-20).

#### **6.3.3.6 Growth rates of *T. thermophilus* expressing AfEst2, ApADH or the synthetic operon.**

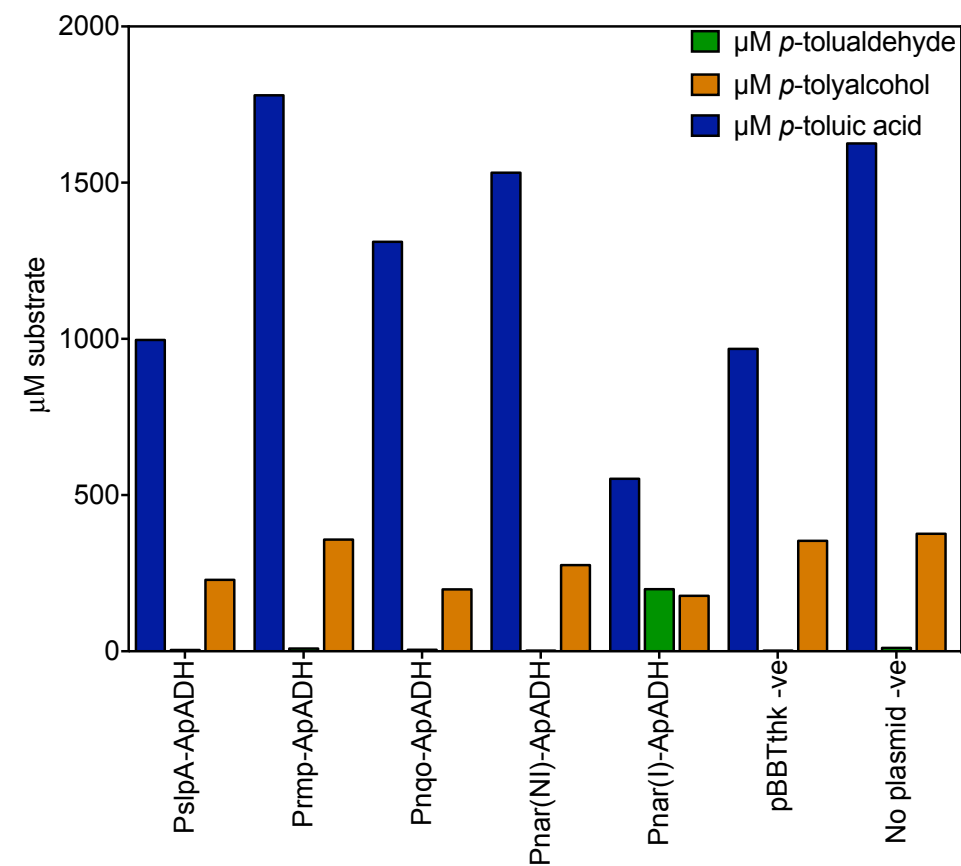
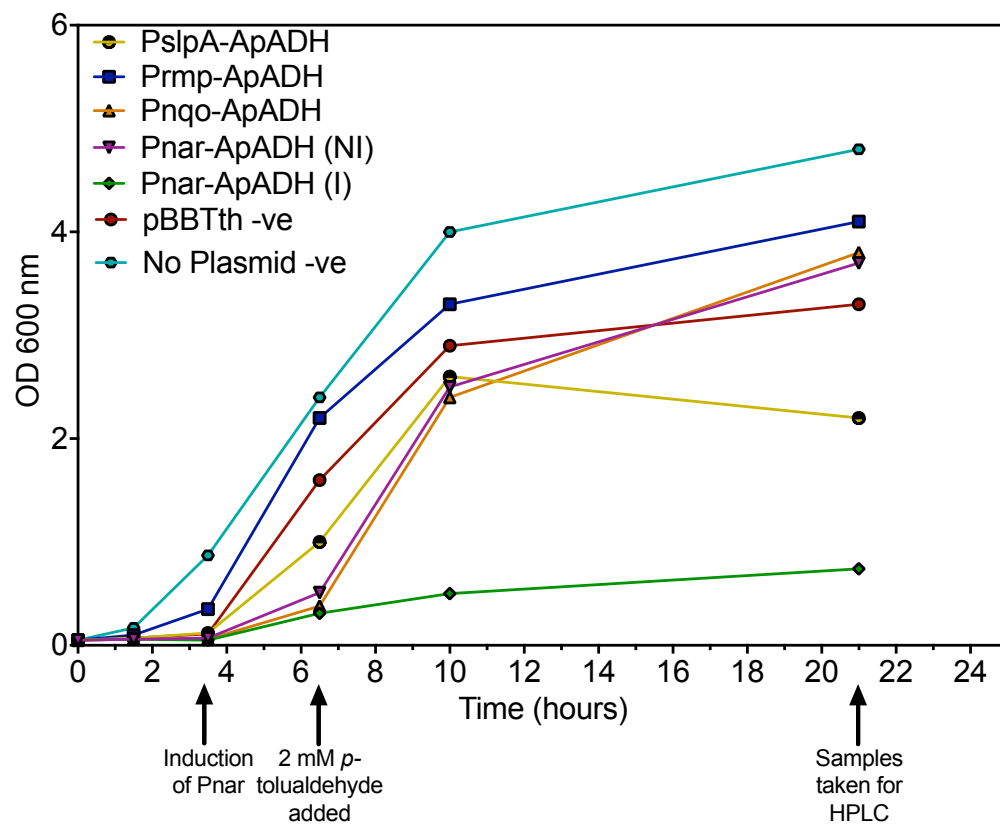
Induction of the Pnar promoter by switching to anaerobic conditions with the addition of 40 mM potassium nitrite stalled growth nearly completely in all cases (Figure 6-17, Figure 6-18 and Figure 6-20). *T. thermophilus* without the addition of the pBBTth plasmid and not challenged by kanamycin was generally able to grow more quickly and to a higher cell density than the other cultures. Expression of AfEst2 or the synthetic operon from the strongest promoter PslpA resulted in much slower growth and lower final cell density. In contrast the PslpA-ApADH culture did not show this to the same degree. This was likely due to the fact that there was no protein expression, although the final cell density was still lower. Expression from other constructs was comparable to empty pBBTTh in most cases.



**Figure 6-17 - Growth curves of *T. thermophilus* constructs transformed with AfEst2 constructs (left), and their conversion of methyl *p*-toluate into *p*-toluic acid (right).**

**Left:** Growth curves of *T. thermophilus* transformed with pBBTth containing various constructs from Figure 6-16. Two Pnar constructs representing a culture that was induced (I), and one which was not (NI). pBBTth –ve refers to *T. thermophilus* transformed with an empty pBBTth plasmid. No plasmid control refers to *T. thermophilus* with no plasmid, and no kanamycin in the growth media. The points at which Pnar was induced, 2 mM methyl *p*-toluate was added, and when samples were taken for analysis, are shown.

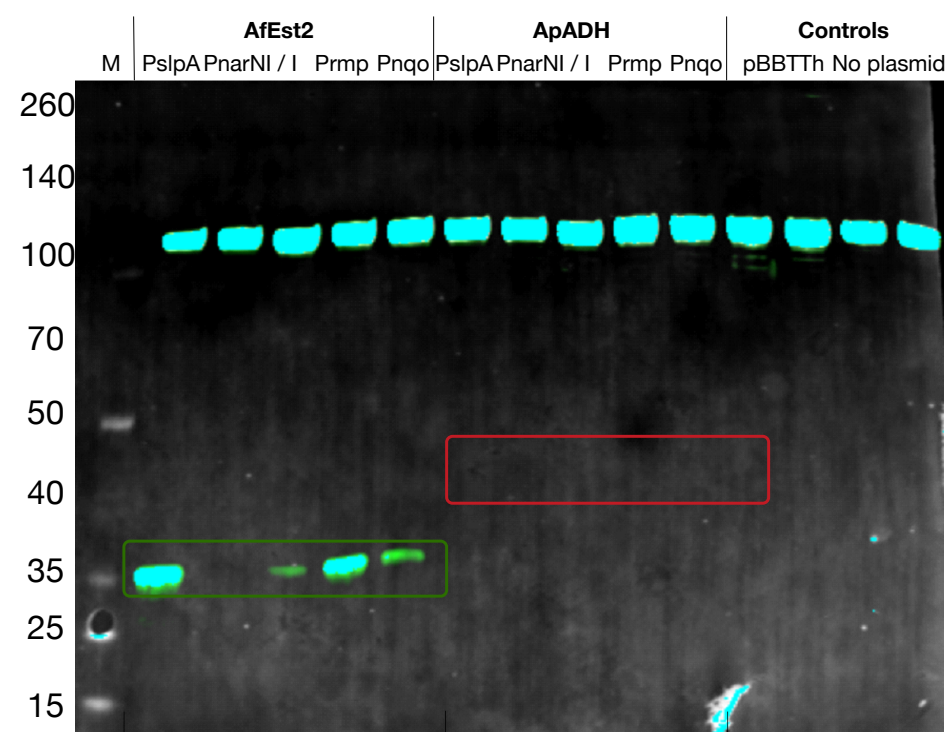
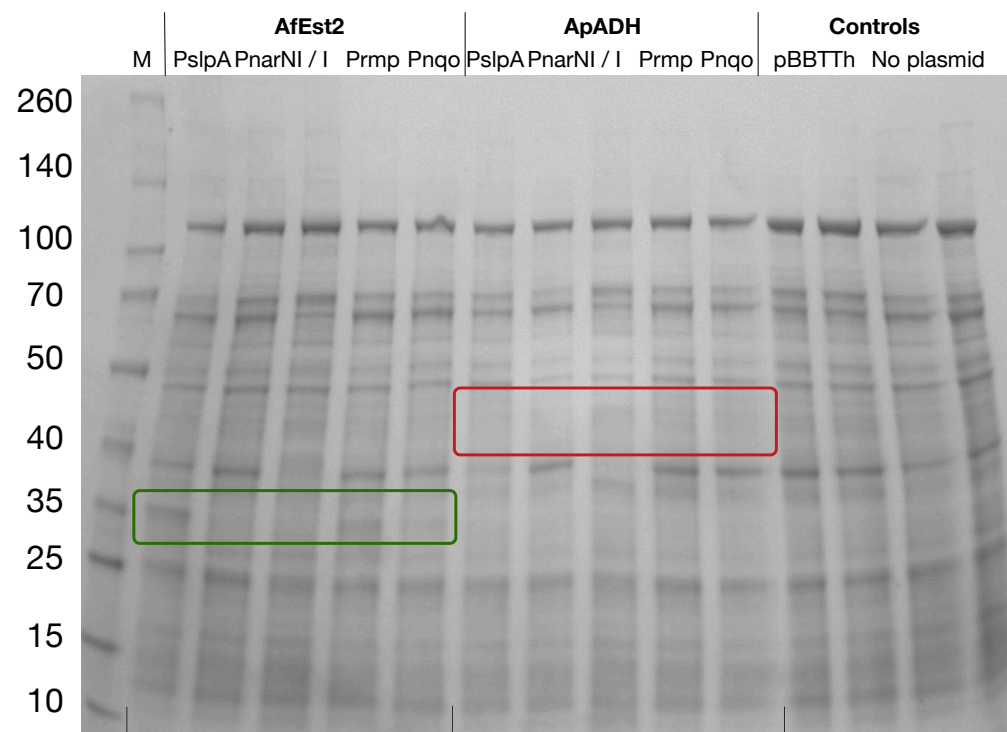
**Right:** Conversion of methyl *p*-toluate into *p*-toluic acid by *T. thermophilus* cells. Concentrations were calculated by reading HPLC peak areas and using standard curves.



**Figure 6-18 - Growth curves of *T. thermophilus* constructs transformed with ApADH constructs (left), and their conversion of *p*-tolualdehyde (right).**

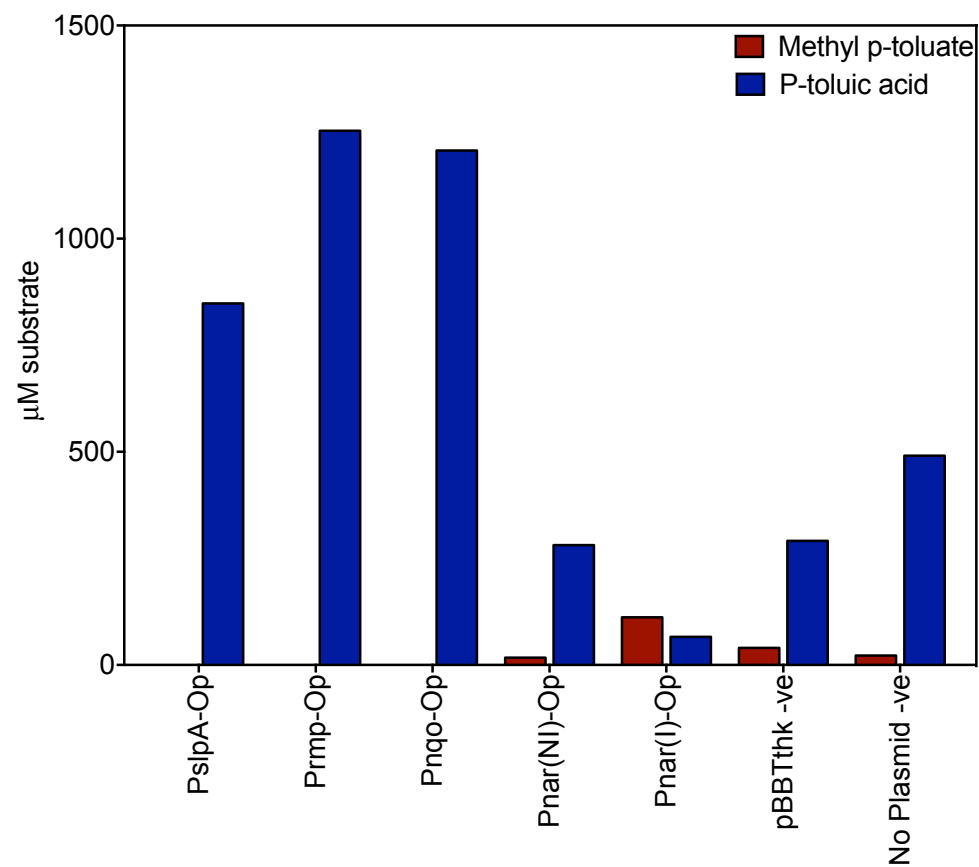
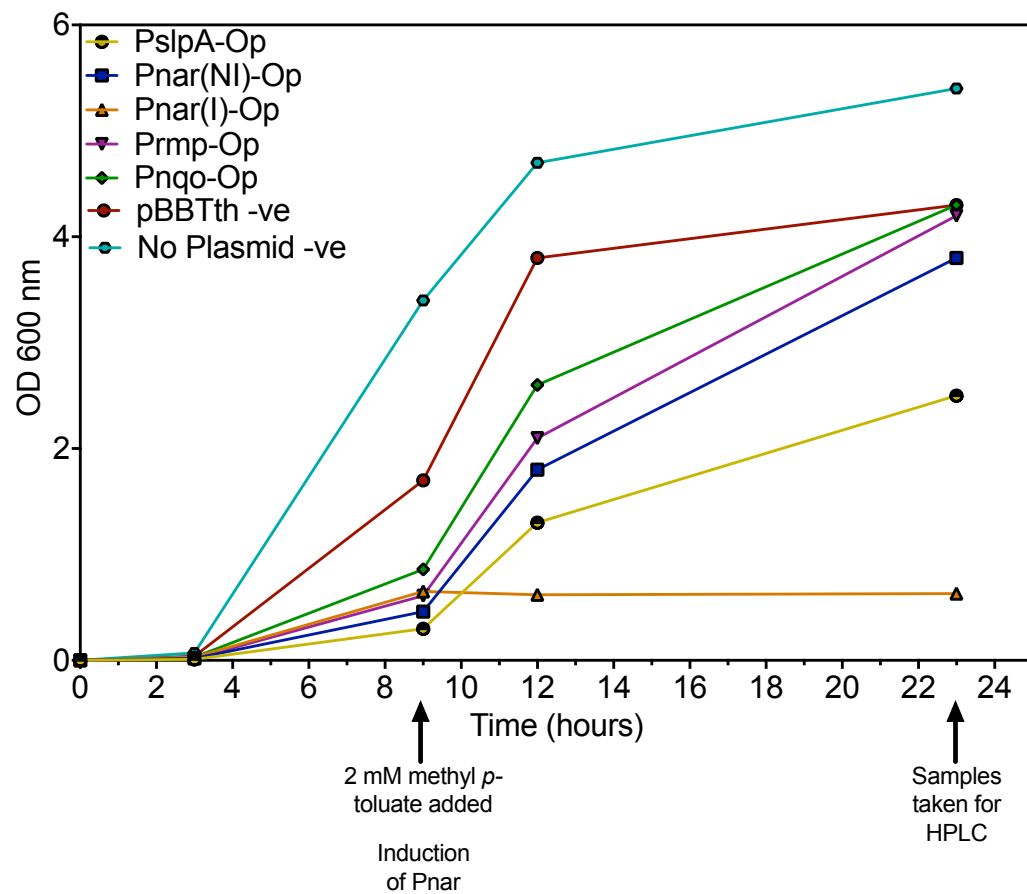
**Left:** Growth curves of *T. thermophilus* transformed with pBBTth containing various constructs from Figure 6-16. Two Pnar constructs representing a culture that was induced (I), and one which was not (NI). pBBTth –ve refers to *T. thermophilus* transformed with an empty pBBTth plasmid. No plasmid control refers to *T. thermophilus* with no plasmid, and no kanamycin in the growth media. The points at which Pnar was induced, 2 mM *p*-tolualdehyde was added, and when samples were taken for analysis, as shown.

**Right:** Conversion of *p*-tolualdehyde into *p*-toluic acid or *p*-tolyl alcohol by *T. thermophilus* cells. Concentrations were calculated by reading HPLC peak areas and using standard curves.



**Figure 6-19 - SDS-PAGE and western blot on *T. thermophilus* total cell lysate, transformed with constructs for the expression of AfEst2 or ApADH only.**

SDS-PAGE and Western blot analysis of the total cell lysate from *T. thermophilus* expressing ApADH (myc tagged) or codon optimised AfEst (FLAG tagged), under the control of PslpA, Pnar, Prmp and Pnqo. Two cultures for Pnar were included, one which was induced (I), and one which was not (NI). Expected sizes of AfEst and ApADH are shown by the green and red boxes respectively.

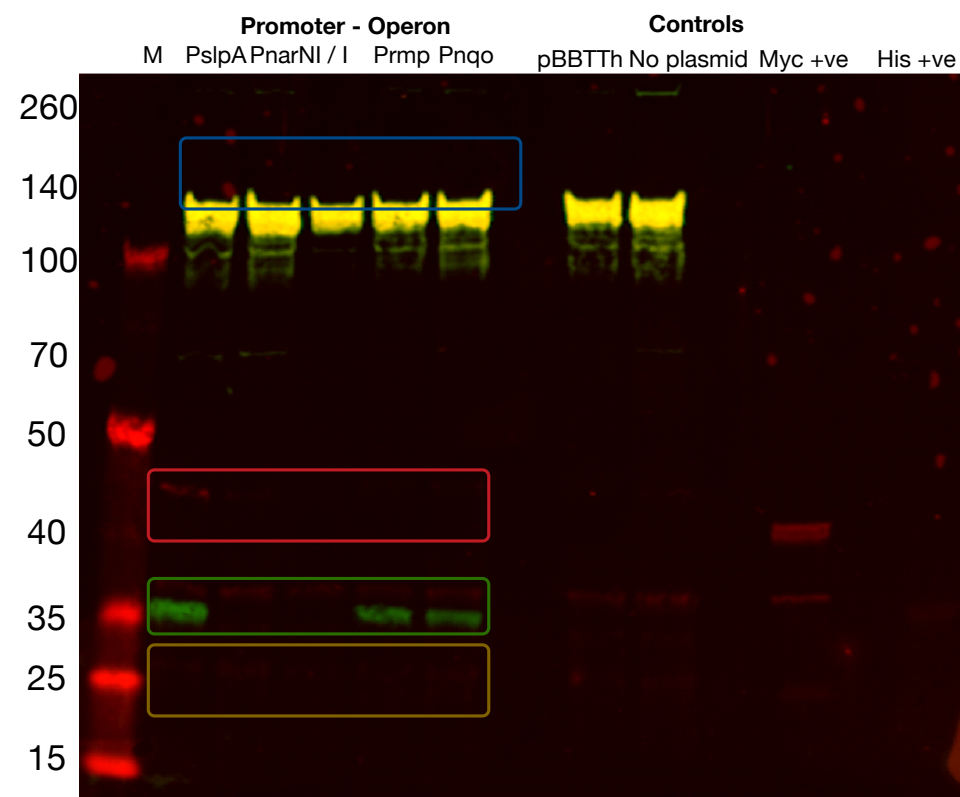
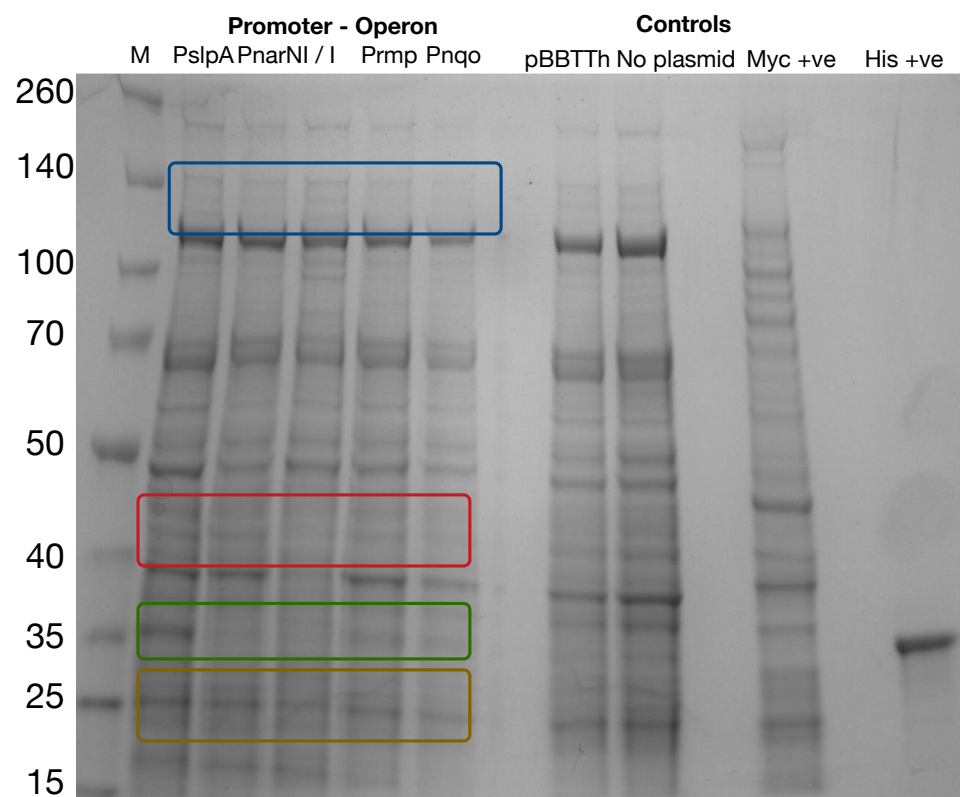




**Figure 6-20 - Growth curves of *T. thermophilus* constructs transformed with synthetic operon constructs (left), and their conversion of methyl *p*-toluate (right).**

**Left:** Growth curves of *T. thermophilus* transformed with pBBTth containing various constructs from Figure 6-16. Two Pnar constructs representing a culture that was induced (I), and one which was not (NI). pBBTth –ve refers to *T. thermophilus* transformed with an empty pBBTth plasmid. No plasmid control refers to *T. thermophilus* with no plasmid, and no kanamycin in the growth media. The points at which Pnar was induced, 2 mM methyl *p*-toluate was added, and when samples were taken for analysis, are shown.

**Right:** Conversion of *p*-tolualdehyde into *p*-toluic acid or *p*-tolyl alcohol by *T. thermophilus* cells. Concentrations were calculated by reading HPLC peak areas and referring to standard curves.



**Figure 6-21 - SDS-PAGE and Western blot of *T. thermophilus* total cell lysate, transformed with the constructs for the expression of a synthetic operon containing mpCAR, AfEst, ApADH and CtTran**

SDS-PAGE and western blot of total cell lysate from *T. thermophilus* expressing a synthetic operon for the expression of mpCAR (6x-His tagged), AfEst2 (FLAG tagged), ApADH (myc tagged) and CtTran (6x-His tagged), under the control of PslpA, Pnar, Prmp and Pnqo. Two cultures for Pnar were included, one which was induced by anaerobic conditions and 40 mM potassium nitrite (I), and one which was not (NI). *T. thermophilus* containing empty pBBTthK and without a plasmid were included as negative controls. Expected sizes of mpCAR, AfEst, ApADH and CtTran are shown by the blue, green, red and yellow boxes respectively.

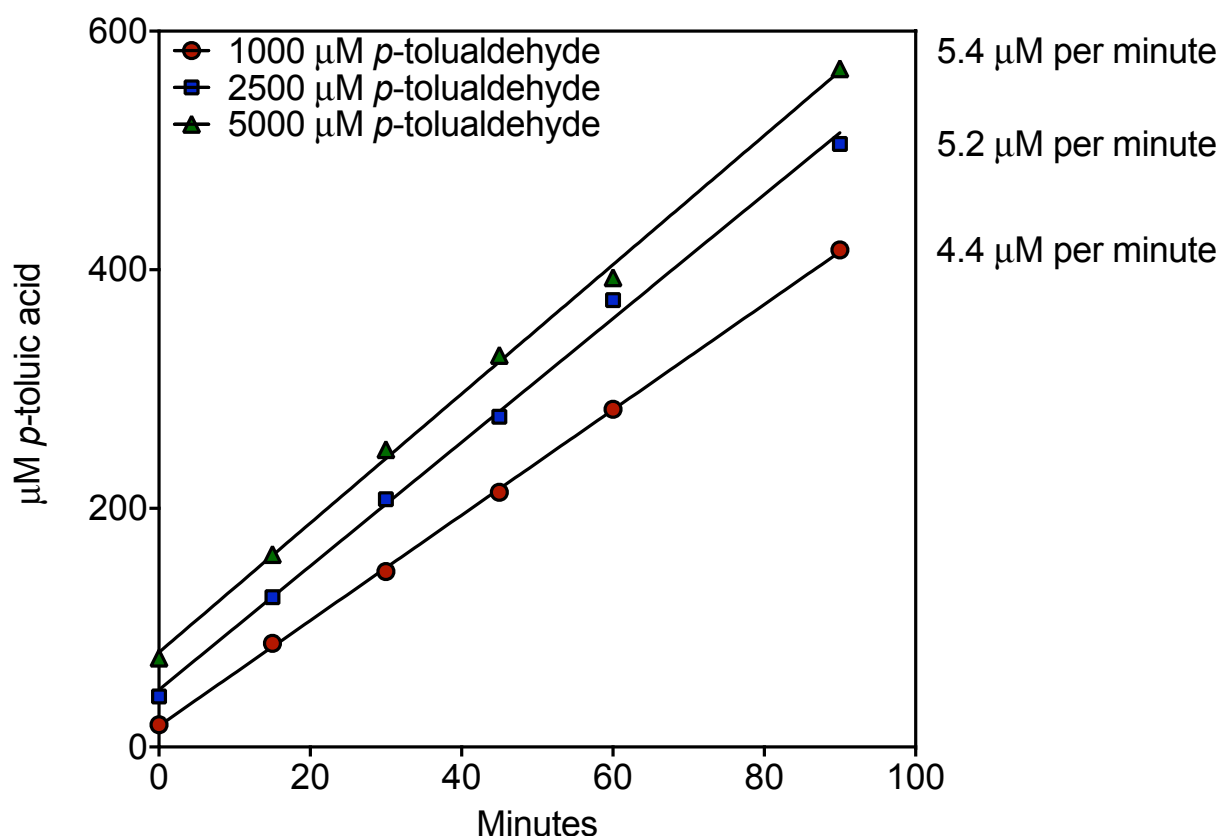
### **6.3.5 Conversion of *p*-tolualdehyde into *p*-toluic acid by native *T. thermophilus* enzymes**

#### **6.3.5.1 Three aldehyde ferredoxin oxidoreductases (AFOs) in the *T. thermophilus* genome**

Conversion of *p*-tolualdehyde into *p*-toluic acid by *T. thermophilus* is likely due to one or more AFO enzymes found in *T. thermophilus*. These enzymes have been shown to carry out the oxidation of aromatic aldehydes into their corresponding acids, with oxidized ferredoxin as an electron acceptor (Heider, Ma and Adams, 1995). There are three *afo* genes in the *T. thermophilus* genome, showing good similarity to well-studied *afo* genes from *Pyrococcus furiosus* (Chan *et al.*, 1995; Heider, Ma and Adams, 1995; Van Den Ban *et al.*, 1999; Ni *et al.*, 2013).

#### **6.3.5.2 Testing the rate of *p*-tolualdehyde oxidation by *T. thermophilus*.**

*T. thermophilus* was grown to plateau phase at approximately OD<sub>600nm</sub> 4.0 and the rate of *p*-tolualdehyde oxidation by these cells measured. Three *p*-tolualdehyde concentrations were tested: 1000, 2500 and 5000 µM. The rate of *p*-toluic acid production was measured by taking samples every 15 minutes for analysis by HPLC. Oxidation of *p*-tolualdehyde was relatively fast in all samples. The presence of these enzymes will be problematic in using *T. thermophilus* for the reduction of *p*-toluic acid or other aromatic compounds. Strategies to delete these enzymes should be considered if the use of a CAR in *T. thermophilus* is to be successful.



**Figure 6-22 – Oxidation of *p*-tolualdehyde by *T. thermophilus* at  $OD_{600nm}$  4.0**

*T. thermophilus* was grown to  $OD_{600nm}$  4.0 before the addition of a *p*-tolualdehyde between 1000 and 5000  $\mu M$ . Samples were taken every 15 minutes into an equal volume of acetonitrile, followed by centrifugation to remove cellular debris. Samples were analyzed by HPLC and quantified against standard curves of *p*-toluic acid and *p*-tolualdehyde.

### **6.3.7 Improving ADH expression in *T. thermophilus*, and the expression of a $\gamma$ -lactamase.**

#### **6.3.7.1 Codon optimisation of ApADH and redesign of 5' CDS for an improved RBS**

ApADH was codon optimized to match the *T. thermophilus* codon usage, and a new RBS designed using the RBS calculator (Table 6-4) (Salis, Mirsky and Voigt, 2009). Due to the specific ApADH sequence, codon optimization was limited by the capabilities of gene synthesis for high GC sequences (IDT was used), and so some compromises had to be made. Also the RBS strength designed was very low, possibly inhibiting successful protein expression. The 5' of the ApADH sequence and RBS was redesigned with the inclusion of an N-terminal myc tag, allowing a much stronger RBS to be designed (Table 6-4). The PslpA promoter and a construct with no promoter were used to test protein expression in *T. thermophilus* (Figure 6-23).

#### **6.3.7.2 Constructs for the expression of an ADH from *Thermus* sp. ATN1, and a $\gamma$ -lactamase from *Sulfolobus solfataricus***

An ADH from *Thermus* sp. ATN1 has been shown to turnover benzaldehyde into benzyl alcohol (Höllrigl *et al.*, 2008), and it was assumed it would therefore show good activity with *p*-tolualdehyde. As this enzyme is from a related *Thermus* species it was also hoped it would show better expression than ApADH. A  $\gamma$ -lactamase from *Sulfolobus solfataricus* (ss- $\gamma$ -lactamase) (Toogood *et al.*, 2004; Hickey *et al.*, 2009), was also identified as an interesting target for expression in *T. thermophilus*, possibly replacing AfEst2 in the cascade reaction and catalyzing a reaction of industrial interest to GSK. Codon optimized C-terminally FLAG tagged sequences were designed with custom RBSs, and cloned under the control of a PslpA or Prmp promoters, as well as without a promoter, to test expression in *T. thermophilus* (Figure 6-23).

#### **6.3.7.3 Expression of codon optimised ApADH, ATN1-ADH and ss- $\gamma$ -lactamase**

*T. thermophilus* cells were transformed with constructs for the expression of ApADH, ATN1-ADH and ss- $\gamma$ -lactamase in pBBTth (Figure 6-23). Overnight cultures were

grown at 70 °C, 200 RPM in TB media with 30 µg / ml kanamycin, following which samples taken for SDS-PAGE and western blot analysis. Expression could not be detected by SDS-PAGE, but all proteins were detected by western blot (Figure 6-24). The western blot was probed using chicken anti-flag and mouse anti-myc primary antibodies. Two secondary antibodies were used, rabbit anti-mouse IRDye 680 and goat anti-chicken IRDye 800CW. Positive controls for myc and FLAG were included, showing up as a strong red or green band respectively. *T. thermophilus* containing an empty pBBTth plasmid was included as a negative control, and non-specific binding of multiple antibodies to a protein at approximately 100 kDa was detected in all *T. thermophilus* samples.

As ApADH showed good expression, and preliminary data suggested ATN1-ADH did not perform any better than ApADH for conversion of *p*-tolualdehyde into *p*-tolyl alcohol. The use of ATN1-ADH in *T. thermophilus* was therefore not investigated further.

#### **6.3.7.4 Testing *p*-tolualdehyde and *p*-toluic acid conversion by *T. thermophilus* cells expressing ApADH.**

It was hoped that the expression of ApADH in *T. thermophilus* would improve the conversion of *p*-tolualdehyde into *p*-tolyl alcohol, and even that *p*-toluic acid might be converted to *p*-tolyl alcohol if native AFOs showed any reductive activity. The equilibrium catalyzed by the native AFO and ADH enzymes will be dependent on the oxidative state of the ferredoxin and NAD(P)H pools, which may be influenced by the growth phase of *T. thermophilus*.

For this reason the growth of the three *T. thermophilus* cultures tested (ApADH with no promoter (NP), PslpA-ApADH and an empty pBBTth –ve control) were synchronized by dilutions to a similar OD<sub>600nm</sub>, followed by growth at 70 °C 200 RPM before another dilution to OD<sub>600nm</sub> 0.1 (shown as hour 0 - Figure 6-25). Growth was followed at OD<sub>600nm</sub>, and was fairly consistent between the cultures.

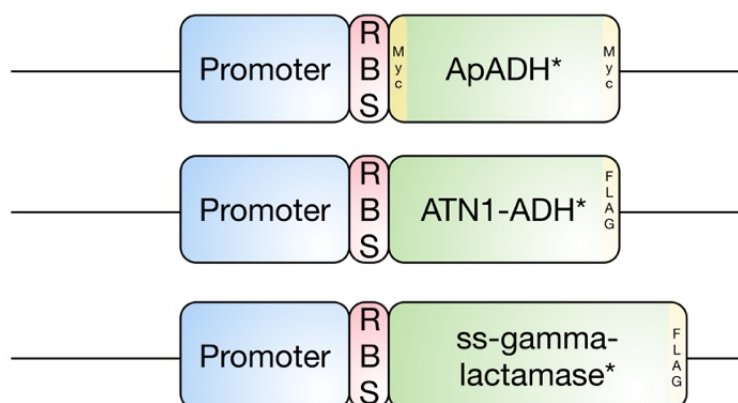
At 4.5 hours, 2 mM *p*-toluic acid (Figure 6-25-A) or *p*-tolualdehyde (Figure 6-25-B) was added. *p*-toluic acid concentration remained fairly constant at 2 mM, with very small amounts of *p*-tolualdehyde and *p*-tolyl alcohol detected. The *p*-tolualdehyde

concentration very quickly decreased to around 500  $\mu$ M after only one hour, and was almost completely converted to *p*-toluic acid 7.5 hours after its addition.

#### **6.3.7.5 Preliminary test for conversion of *p*-toluamide by *T. thermophilus* cells expressing ss- $\gamma$ -lactamase.**

Overnight cultures of *T. thermophilus* containing Prmp-ss- $\gamma$ -lactamase, ss- $\gamma$ -lactamase only and an empty pBBTth –ve control were set up in TB media + 30  $\mu$ g / ml kanamycin at 70 °C 200 RPM. 2 mM *p*-toluamide was added and the cultures incubated for a further 8 hours before samples were taken for analysis by HPLC. Between 260 and 290 mM *p*-toluic acid was detected in all the *T. thermophilus* cultures expressing ss- $\gamma$ -lactamase, with very low levels found in the empty pBBTth –ve control (Figure 6-26).





**Figure 6-23 – New constructs for the expression of codon optimized ApADH, ATN1-AND and ss- $\gamma$ -lactamase**

From RBS-CDS parts, constructs were made for the expression of ApADH, ATN1-ADH or ss- $\gamma$ -lactamase. \* is used to indicate a codon optimized sequence.

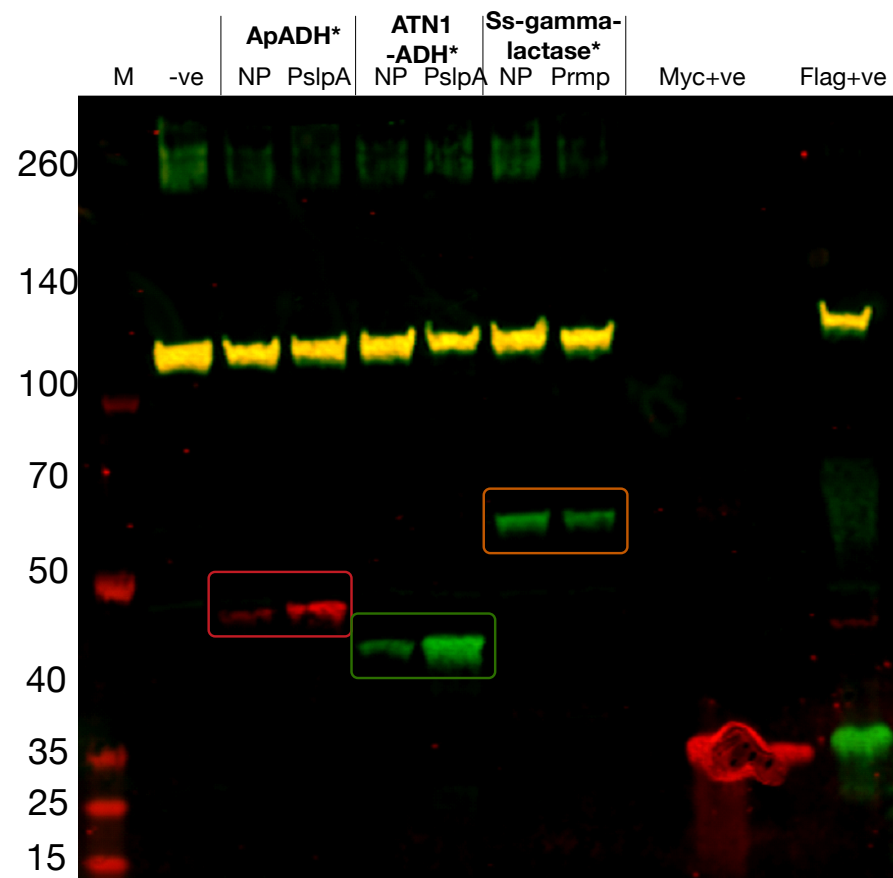
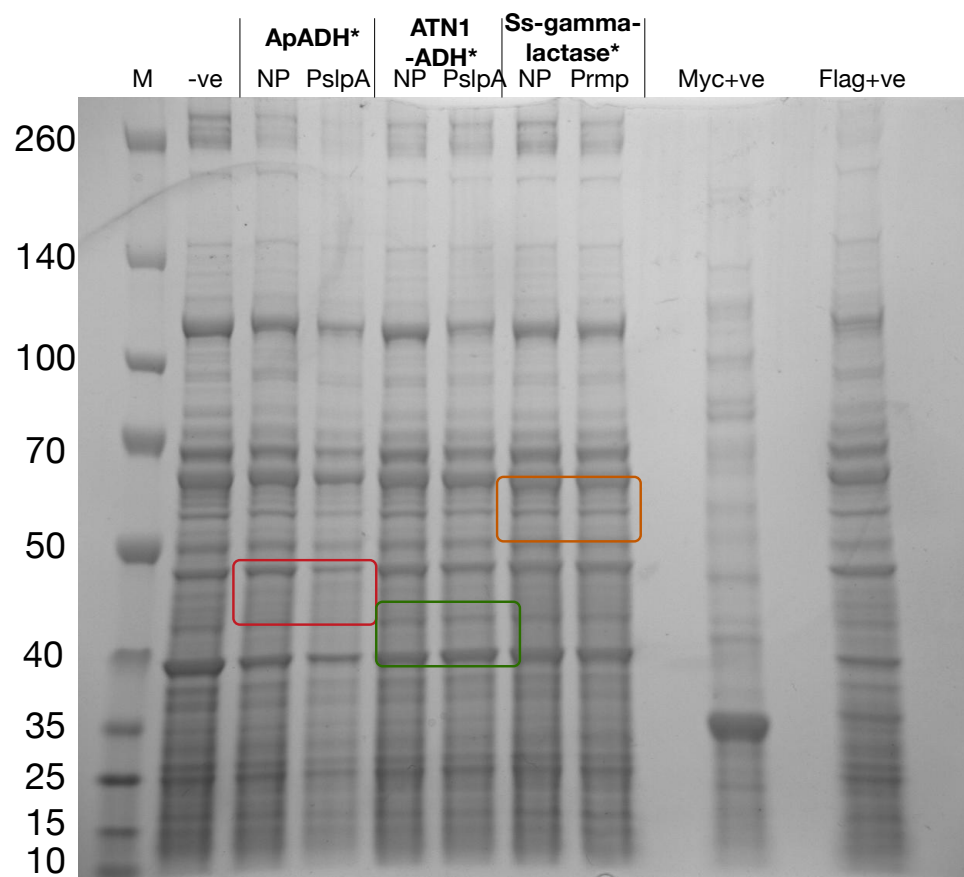
Promoter: A promoter part, either PslpA, Pnar, Prmp, Pnqo or no promoter relying on read through on plasmid.

RBS: A custom RBS for each CDS, designed using the RBS calculator.

ApADH\*: Codon optimized gene for the expression of ApADH, a C- and N-terminal Myc tag was included.

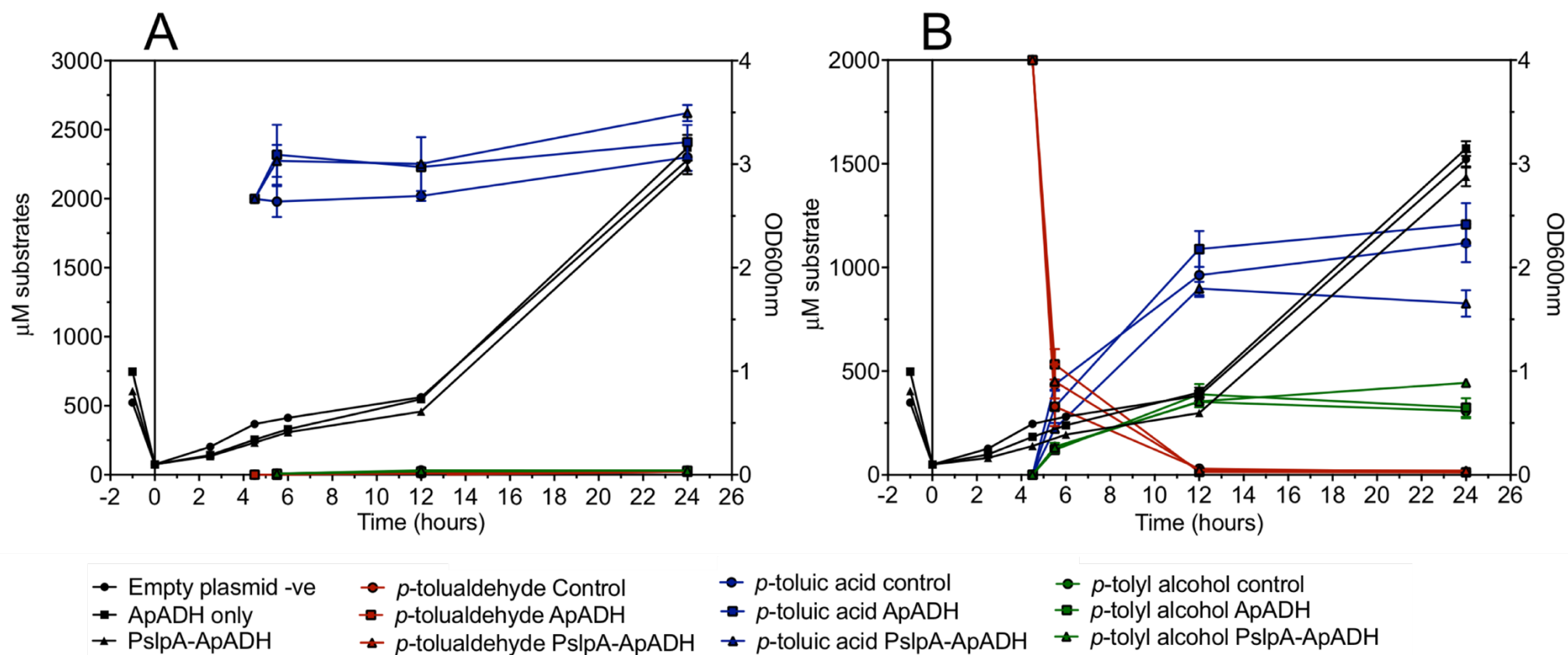
ATN1-ADH\*: Codon optimized gene for the expression of ATN1-ADH, a C-terminal FLAG tag was included.

ss- $\gamma$ -lactamase\*: Codon ptimized gene for the expression ss- $\gamma$ -lactamase, a C-terminal FLAG tag was included.



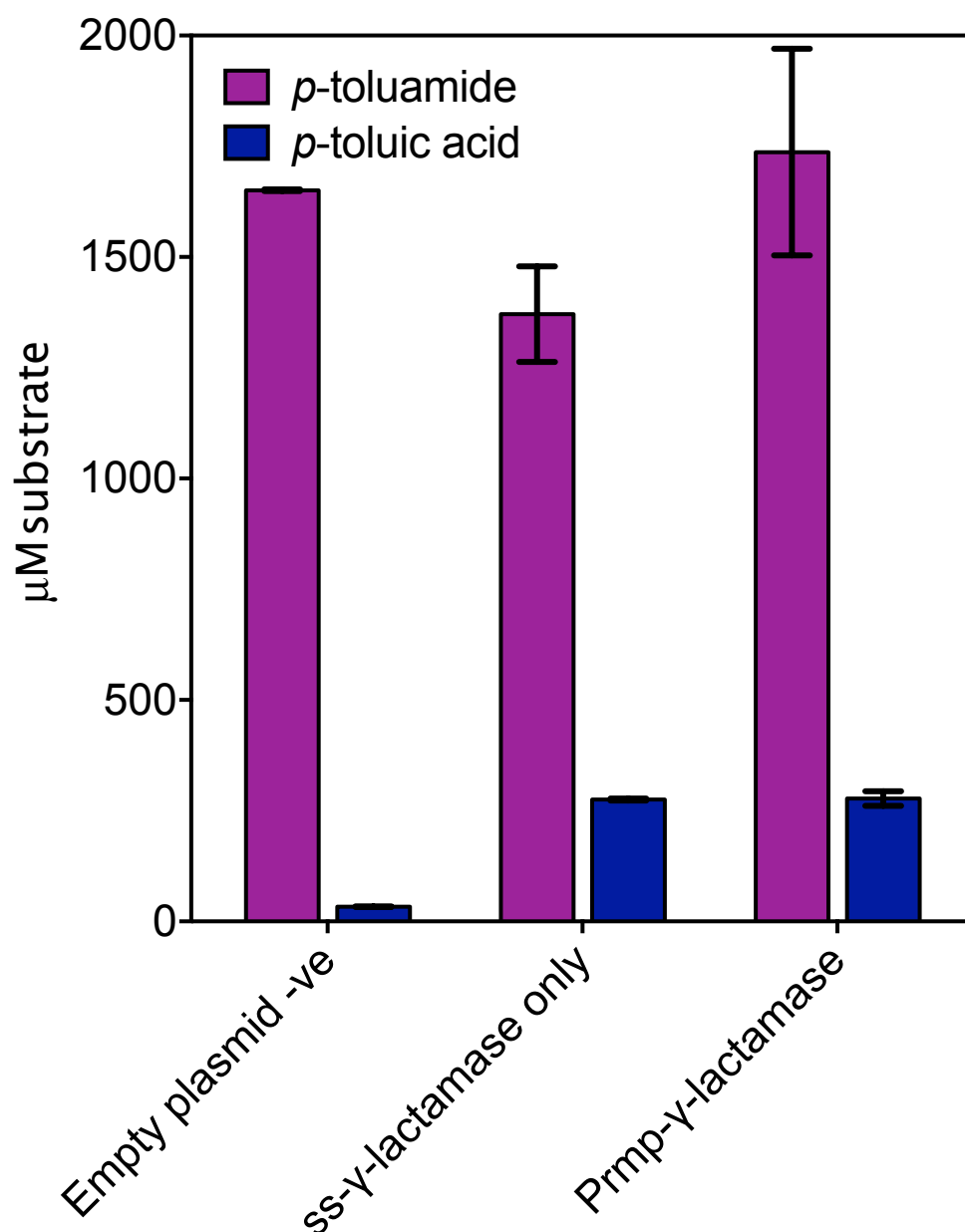
**Figure 6-24 - SDS-PAGE and western blot analysis of *T. thermophilus* total cell lysate, transformed with constructs for the expression of codon optimized ApADH, ATN1-ADH and ss- $\gamma$ -lactamase.**

SDS-PAGE and western blot analysis of total cell lysate from *T. thermophilus* expressing codon optimised ApADH (N- and C-terminally myc tagged), codon optimised ATN1-ADH (FLAG tagged), or codon optimised ss- $\gamma$ -lactamase (FLAG tagged). Proteins were expressed under the control of the PslpA promoter, or without a promoter (NP). Expected sizes for ApADH, ATN1-ADH and ss- $\gamma$ -lactamase are shown by the red, green and orange boxes respectively.



**Figure 6-25 - Conversion of *p*-toluic acid or *p*-tolualdehyde by *T. thermophilus* cells expressing ApADH**

The optical density of *T. thermophilus* cultures and concentrations of *p*-toluic acid, *p*-tolualdehyde and *p*-tolyl alcohol are shown. **A.** shows the addition of 2 mM *p*-toluic acid to growing *T. thermophilus* cells. **B.** shows the addition of 2 mM *p*-tolualdehyde to growing *T. thermophilus* cells. OD<sub>600nm</sub> is shown in black, and is on the right hand axis of both A and B. *p*-toluic acid is shown in blue, *p*-tolualdehyde in red and *p*-tolyl alcohol in green. *T. thermophilus* cultures transformed with an empty pBBTth is shown by the circles, containing only ApADH by the squares and ApADH under the control of PslpA by triangles. Error bars so the standard deviation of three readings.



**Figure 6-26 – Conversion of *p*-toluamide by *T. thermophilus* cells expressing ss-γ-lactamase.**

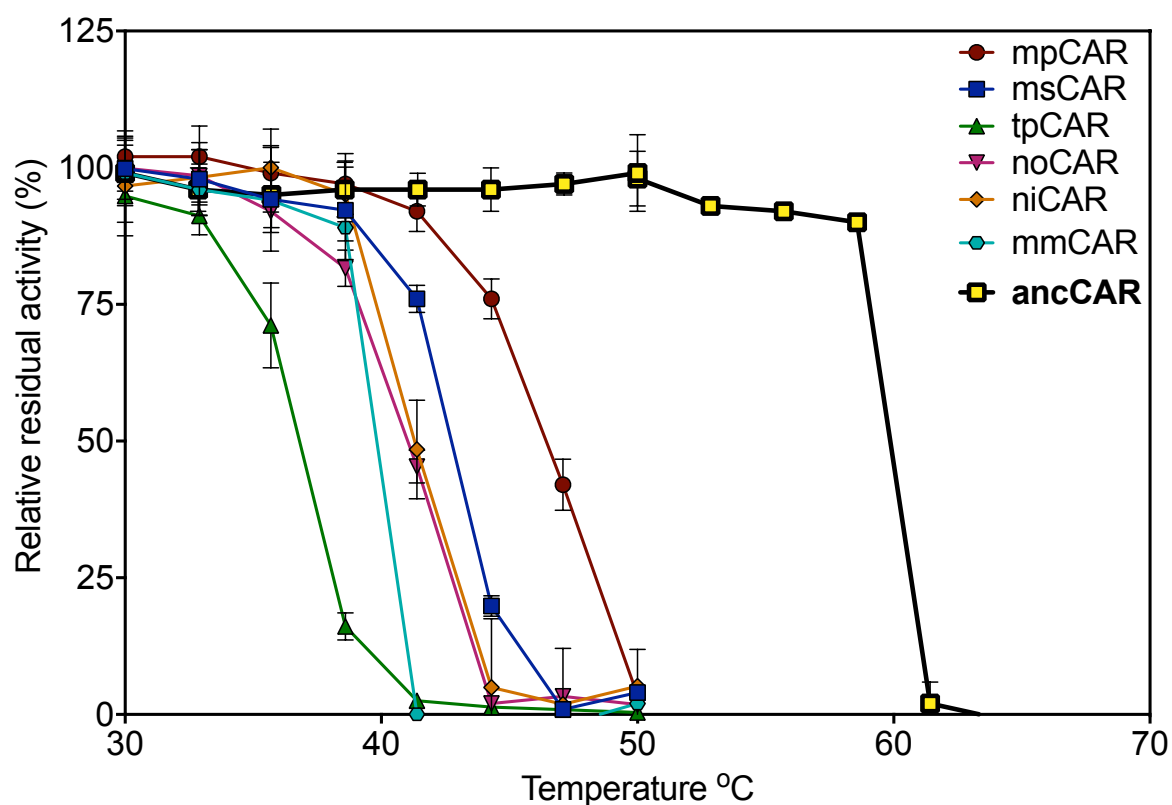
Concentrations of *p*-toluamide (purple) and *p*-toluic acid (blue) are shown following incubation of 2 mM *p*-toluamide with cultures of *T. thermophilus* expressing ss-γ-lactamase. *T. thermophilus* containing an empty pBBTth vector was used as the control. *T. thermophilus* expressing ss-γ-lactamase was able to transform between 265 and 290 9μM *p*-toluamide into *p*-toluic acid, while the control showed only a very low concentration of *p*-toluic acid. Error bars show the standard deviation of two repeats.

### 6.3.8 Improving CAR expression in *T. thermophilus*

#### 6.3.8.1 Ancestral gene reconstruction for a more thermostable CAR

A thermostable CAR is necessary for expression in *T. thermophilus*, but to date no thermostable CARs have been identified. Efforts to find a CAR from a thermophilic source were also unsuccessful (data not shown). mpCAR, as discussed in chapter 4, comes from the moderate thermophile *Mycobacterium phlei* and is stable up to approximately 47 °C *in vitro* (Figure 6-27). It was hoped *in vivo* stability might be greater, allowing some CAR activity at lower *T. thermophilus* growth temperatures. However ideally a more thermostable CAR would be used for expression in *T. thermophilus*.

Adam Thomas (University of Exeter) investigated the use of ancestral sequence reconstruction to produce a more thermostable CAR (ancCAR), which was stable up to 59 °C (Figure 6-27).



**Figure 6-27 - Thermostability of a number CAR proteins in comparison to ancCAR**

The residual activity after incubation for half an hour at a range of temperatures, relative to a control kept on ice, is shown. mpCAR: CAR from *Mycobacterium phlei*, msCAR: CAR from *M. smegmatis*, tpCAR: CAR from *Tsukamurella paurometabola*, noCAR: CAR from *Nocardia otitidiscaviarum*, niCAR: *N. iowensis*, mmCAR: *M. marinum*, ancCAR: Ancestrally reconstructed CAR.



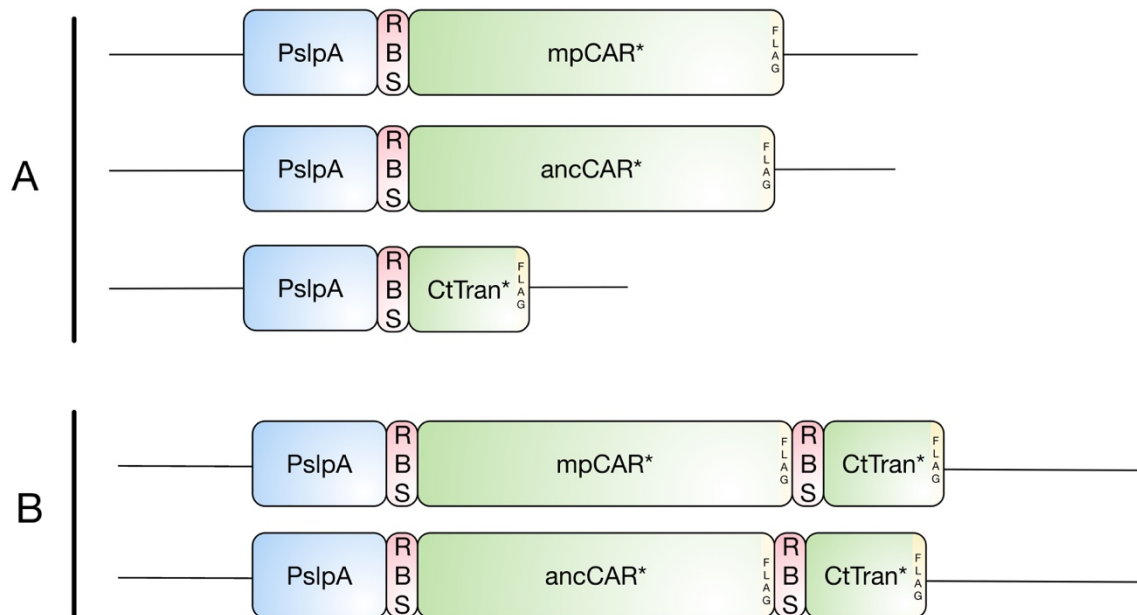
#### 6.3.8.2 Design and expression of codon optimised ancCAR, mpCAR, and a re-optimised CtTran.

Codon optimized sequences for the expression of mpCAR (CAR from *Mycobacterium phlei*), ancCAR (Ancestrally reconstructed CAR) and CtTran were designed, featuring N-terminal FLAG tags. RBSs were designed using the RBS calculator set to maximize RBS strength. The PslpA promoter was placed ahead of these genes for their expression in *T. thermophilus* in pBBTth (Figure 6-28-A).

*T. thermophilus* containing these constructs were grown overnight in TB media + 30 µg / ml kanamycin at 70 °C, 200 RPM. The temperature was then dropped to 60 °C for a further 6 hours (given the lower thermostability of the CARs) before samples were taken for SDS-PAGE and Western blot analysis. FLAG tagged protein was detected at the expected molecular weight for all proteins by western blot analysis, but were not detectable by SDS-PAGE (Figure 6-29). The western blot was probed using a chicken anti-flag primary antibody, and a goat anti-chicken IRDye 800CW secondary antibody. A positive FLAG control was used, although was only weakly detected. *T. thermophilus* containing an empty pBBTth plasmid was included as a negative control, and non-specific binding of multiple antibodies to a protein at approximately 100 kDa was detected in all *T. thermophilus* samples.

#### 6.3.8.3 Expression of CAR-CtTran operons for testing CAR activity in *T. thermophilus*

Having shown successful expression of mpCAR, ancCAR and CtTran, CtTran was cloned downstream each of the CARs creating two protein operons (Figure 6-28-B). Expression of the CARs with CtTran was tested as before, with 2 mM *p*-toluic acid added upon reducing temperature to 60 °C. Approximately 6 hours after reducing the temperature to 60 °C samples were taken for western blot analysis and HPLC. Western blot analysis of the cell lysates did not provide conclusive results, although suggests the CARs and CtTran were expressed together (Figure 6-30). No conversion of 2 mM *p*-toluic acid into *p*-tolualdehyde or *p*-tolyl alcohol was detected by these constructs when analyzed by HPLC (data not shown).



**Figure 6-28 – Constructs for the expression of codon optimized mpCAR, ancCAR and CtTran from the PslpA promoter**

From RBS-CDS parts, constructs were made for the expression of **A**: mpCAR, ancCAR or CtTran, or **B**: mpCAR or ancCAR in combination with CtTran. \* is used to indicate a codon optimized sequence.

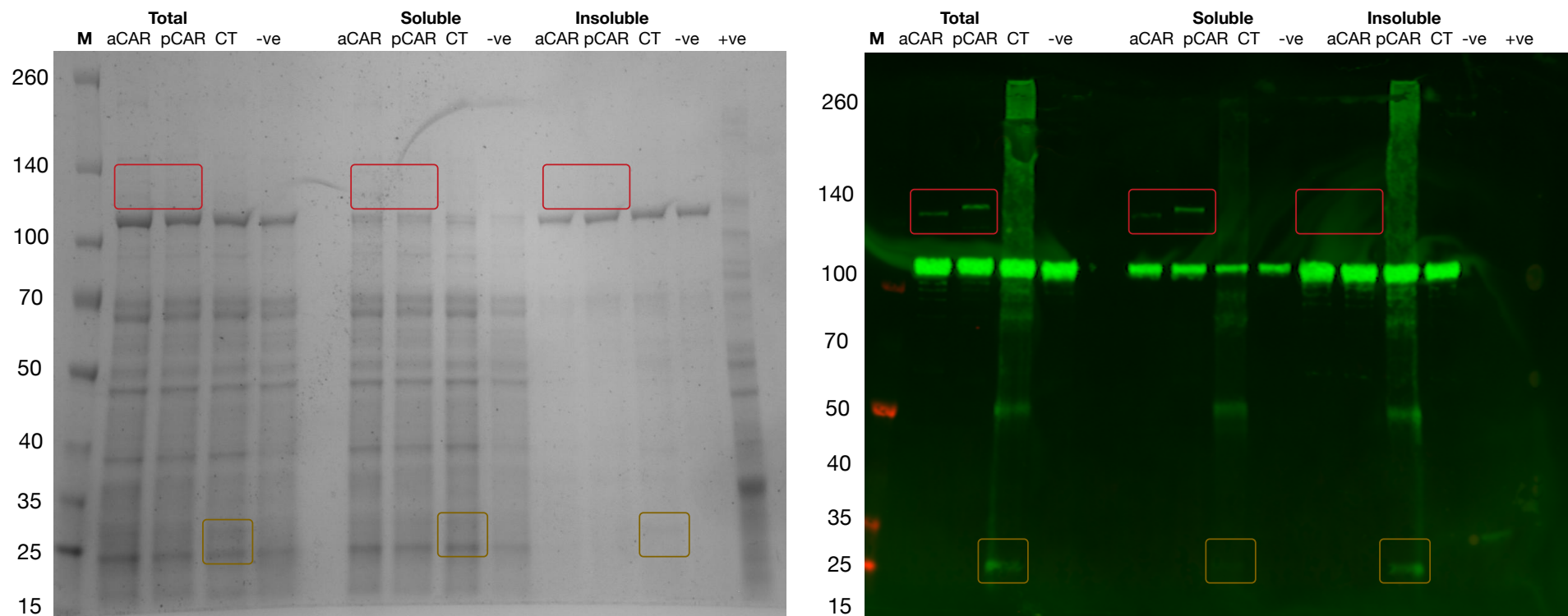
PslpA: The PslpA promoter.

RBS: A custom RBS for each CDS, designed using the RBS calculator.

mpCAR\*: Codon optimized gene for the expression of mpCAR, a C-terminal FLAG tag was included.

ancCAR\*: Codon optimized gene for the expression of ancCAR, a C-terminal FLAG tag was included.

CtTran\*: Codon optimized gene for the expression of CtTran, a C-terminal FLAG tag was included.



**Figure 6-29 – SDS-PAGE and western blot analysis of *T. thermophilus* total cell lysate with soluble and insoluble fractions, transformed with constructs for the expression of codon optimized ancCAR, mpCAR and CtTran only, under the control of the PslpA promoter.**

SDS-PAGE and western blot of total cell lysate from *T. thermophilus* expressing codon optimised mpCAR (FLAG tagged and shown as pCAR), codon optimised ancCAR (FLAG tagged and shown as aCAR), or codon optimised CtTran (FLAG tagged and shown as CT) from the PslpA promoter. Expected sizes for mpCAR and ancCAR are shown by the red boxes, and for CtTran by the yellow boxes.

### **6.3.9 Expression of a Cas9 in *T. thermophilus***

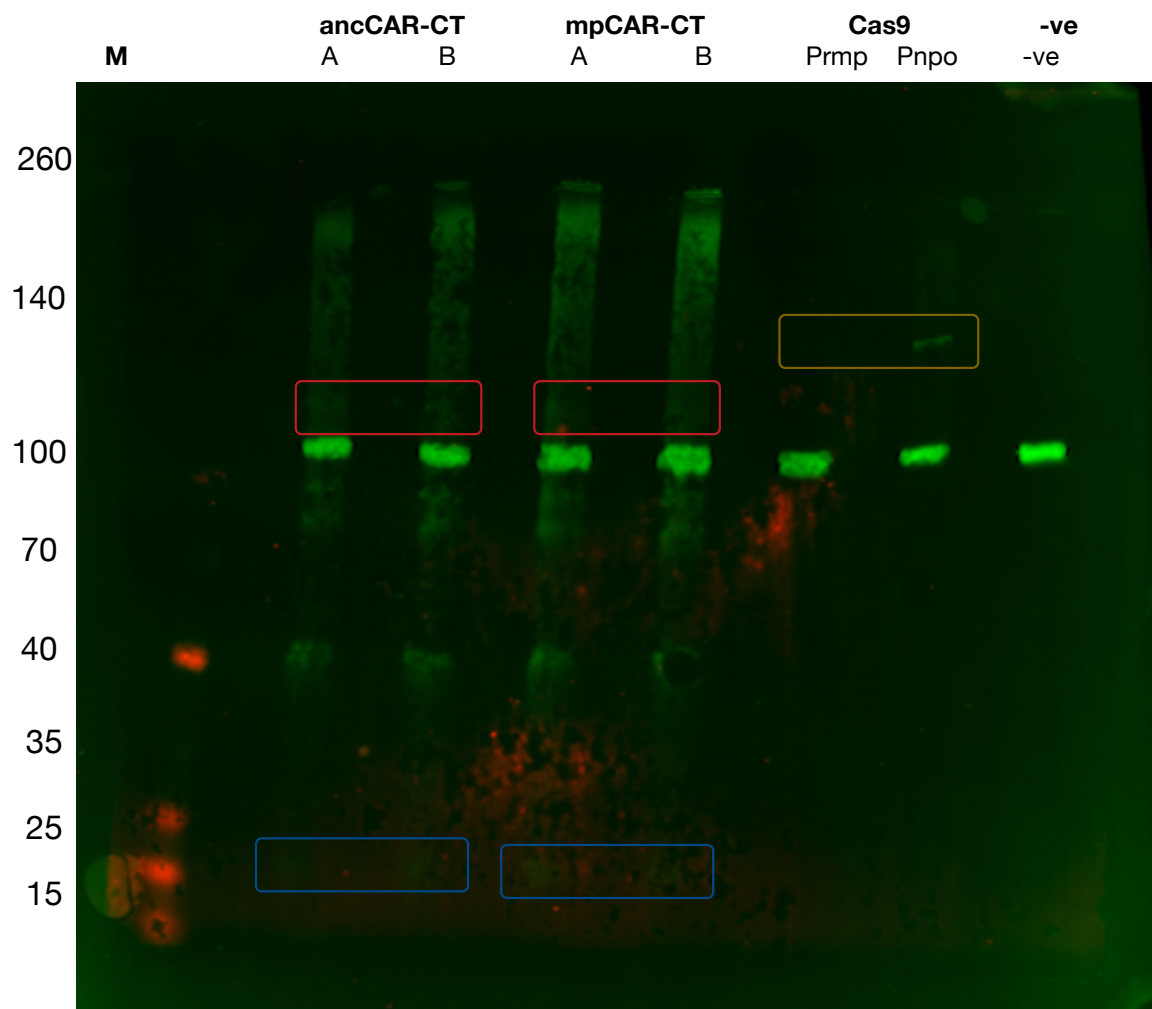
#### **6.3.9.1 Designing constructs for the expression of Cas9 in *T. thermophilus***

A codon optimized and N-terminal FLAG tagged sequences for the expression of the Cas9 protein from *Streptococcus thermophilus* were designed assembled by Gibson assembly.

#### **6.3.9.2 Strong Cas9 expression appears to be toxic to *T. thermophilus***

Cas9 under the control of PslpA, Prmp and Pnqo were transformed into *T. thermophilus*. With expression of Cas9 under the control of PslpA no colonies were found. Expression from Prmp gave two or three colonies, while Pnqo resulted in many colonies. It is clear that high levels of Cas9 expression is toxic to *T. thermophilus* growth.

Overnight cultures containing Prmp and Pnqo Cas9 constructs were grown overnight in TB media + 30 µg / ml kanamycin. The total cell lysate was analyzed by western blot, showing expression of a FLAG tagged protein of the correct molecular weight of Cas9, under the control of Pnqo. No protein was detected from the Prmp construct (Figure 6-30). The western blot was probed using a chicken anti-flag primary antibody, and a goat anti-chicken IRDye 800CW secondary antibody. A positive FLAG control was used, although was only weakly detected. *T. thermophilus* containing an empty pBBTth plasmid was included as a negative control and non-specific binding of multiple antibodies to a protein at approximately 100 kDa was detected in all *T. thermophilus* samples.



**Figure 6-30 – Western blot on *T. thermophilus* total cell lysate, transformed with constructs for the expression of codon optimized ancCAR or mpCAR in combination CtTran (all FLAG tagged). And for expression of a Cas9 protein (FLAG tagged) from *S. thermophilus*.**

SDS-PAGE and western blot of total cell lysate from *T. thermophilus* expressing codon optimised mpCAR or ancCAR in combination with CtTran (all FLAG tagged), under the control of the PslpA promoter, or codon optimised Cas9 from *S. thermophilus* (FLAG tagged) under the control of the Prmp or Pnqo promoters. Expected sizes for mpCAR and ancCAR are shown by the red boxes, for CtTran by the blue boxes and for Cas9 by the yellow box.

### 6.3.10 Summary of Results

This study aimed to develop and test the use of *T. thermophilus* HB27 as a host organism for whole-cell biocatalysis. To facilitate the modular construction of DNA parts, a *T. thermophilus* BioBricks vector was constructed. Using this vector and a sfGFP reporter we were able to characterize a number of constitutive promoters based on their sfGFP expression, providing a panel of promoters with a range of known strengths. The use of the Pnar promoter was also investigated although the requirement for anaerobic conditions limited its usefulness.

The RBS calculator was used to design strong RBSs for all of the proteins expressed in *T. thermophilus* throughout this study. We have demonstrated this to be an effective method for creating custom RBSs for *T. thermophilus* for maximum protein expression.

The pathway shown in Figure 6-5 was to be carried out in *T. thermophilus* to demonstrate its use for whole-cell biocatalysis. However in practice, while all the enzymes could be expressed, only the hydrolysis of methyl *p*-toluate was achieved in *T. thermophilus*, due to the strong oxidation of *p*-tolualdehyde to *p*-toluic acid by native enzymes.

In summary we have demonstrated the use of *T. thermophilus* for whole-cell biocatalysis, but more work is needed to complete the pathway shown in Figure 6-5. We have also started the process of designing and characterizing well-defined parts for the predictable construction of new pathways in *T. thermophilus*.

## 6.4 Discussion

### 6.4.1 Designing new *T. thermophilus* vectors

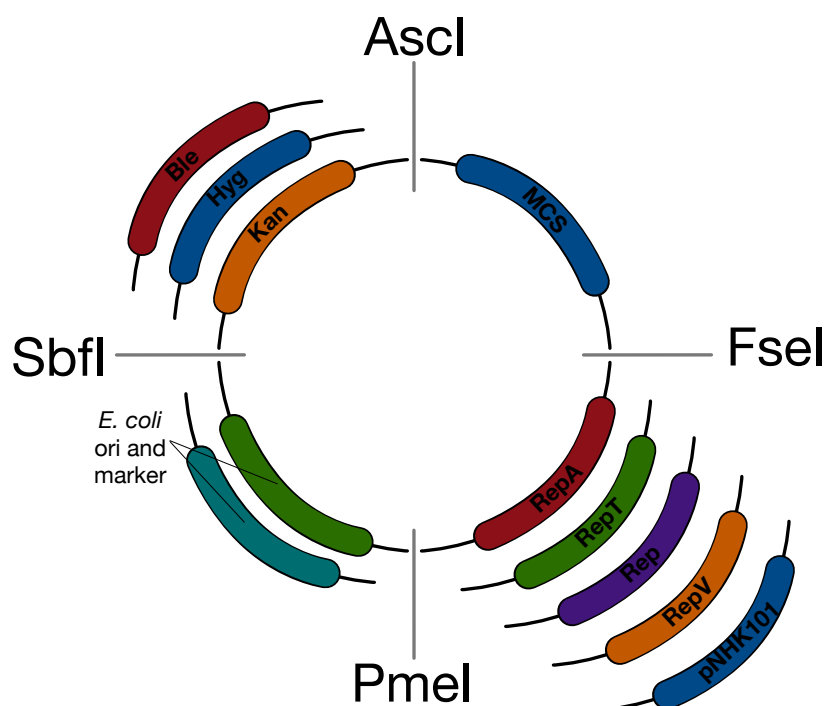
#### 6.4.1.1 Minimisation of vector size inhibited transformation into *T. thermophilus*

The first BioBricks vector featured a minimum of apparently junk DNA but was unable to replicate in *T. thermophilus* (Figure 6-7). Reasons were that either some unknown essential piece of DNA was removed in this design, or a problem with the close proximity of parts such as the replication region for *T. thermophilus* and kanamycin resistance cassette inhibited antibiotic resistance or plasmid replication in *T. thermophilus*. Possibly the inclusion of the antibiotic resistance cassette ahead of the region of replication rather than after it may have offered a better design.

The inclusion of a larger section of DNA taken directly from pMKE2 allowed a second plasmid to replicate in *T. thermophilus* (Figure 6-9), although it is unclear whether this is due to the inclusion of a piece of DNA that was missing previously, or because the parts are better separated from each other.

#### 6.4.1.2 Defined vector parts can be combined in a number of ways for the assembly of new vectors.

Recent work on a modular system for *Clostridium-E. coli* shuttle plasmids allows the combination of many of the components from disparate plasmids to be combined to create new vectors suitable for specific hosts and applications. Using a system of type II restriction enzymes, components can be combined easily as modules, allowing a large number of different vectors to be constructed (Heap *et al.*, 2009). A similar system could be envisaged for the construction of new *T. thermophilus* vectors, using the parts already established and described in the introduction. An example of how much a system might operate is shown in Figure 6-31.



**Figure 6-31 – Hypothetical modular system for the creation of *T. thermophilus* vectors from established parts.**

A modular system for the assembly of DNA parts for the construction of new *T. thermophilus* vectors - inspired from (Heap *et al.*, 2009). A number of regions of replication for *T. thermophilus* and thermostable antibiotic resistance markers are available which could be combined in a number of combination. Different selectable markers, origins of replication for *E. coli* and multiple cloning sites (MCSs) could be used. A system of type II restriction enzymes suggested for the modular construction of *Clostridium-E. coli* shuttle plasmids (Heap *et al.*, 2009), could be used for a similar approach in *T. thermophilus*.

## 6.4.2 Designing genes for expression in *T. thermophilus*

### 6.4.2.1 Codon optimisation was essential for protein expression in *T. thermophilus*

Codon optimization proved to be essential for protein expression in *T. thermophilus*. Initially native genes for mpCAR and ApADH were used and no expression was detected (Figure 6-19 and Figure 6-21). When codon optimized constructs were used protein was detected by western blot analysis (Figure 6-24 and Figure 6-29), however expression was still at a relatively low level compared to the more highly



expressed AfEst2 (Figure 6-19). Clearly codon optimization is important for gene expression in *T. thermophilus*, but the process of optimization itself is in need of optimization. As discussed previously, many other factors can be considered other than codon usage and these may be relevant for optimal expression in *T. thermophilus*.

#### **6.4.2.2 Using the RBS calculator for RBS design for expression in *T. thermophilus***

The RBS calculator was used to design custom RBSs for all the genes in this study, set to maximize RBS strength. Initially RBS BioBrick parts were created to allow modulation of RBS strength through the introduction of new RBS BioBrick parts. However BioBrick cloning results in a scar sequence between the SD and start codon, constraining the design space available to the RBS calculator. This limitation along with the extra time required to clone RBS sites meant it was much more efficient to have RBS sites gene synthesized in front of new genes directly. However this limits their reuse with new RBSs.

A range of RBS strengths were predicted by the calculator for the RBSs designed (Table 6-4). However these do not seem to correlate well with the observed protein expression. Arguably the best protein expression was observed for AfEst2, with a predicted RBS strength of only 69,963 (Figure 6-19). In contrast the RBS designed for ancCAR had a predicted strength of 1,287,767 yet was not detected by SDS-PAGE, and only weakly by western blot analysis (Figure 6-29). Possibly the predictive power of the RBS calculator for *T. thermophilus* is not reliable, or other factors have a greater influence on protein expression. However it is clear that the RBS calculator can be used for the design of functional RBSs for *T. thermophilus*.

### **6.4.3 Promoter characterisation**

#### **6.4.3.1 A range of promoters of varying strength were characterised using sfGFP as a reporter.**

Promoters were successfully characterized in terms of their strength using sfGFP as a reporter (Figure 6-15). As expected PslpA was by far the strongest promoter, useful where high levels of expression are required. However in some cases

promoters can be too strong, for example when the expressed protein is toxic to the cell, or strong expression inhibits reasonable growth rates. In these cases it is useful to have a characterized set of promoters covering a range of strengths. Prmp offers a more moderate level of protein expression for these cases, followed by Pnqo.

Characterization of the promoters would be better achieved looking at sfGFP expression over all phases of growth, as protein expression is dependent on growth phase. For example it is likely that the regulation of PslpA by SlrA and SlpM is not consistent across all phases of growth. Using *E. coli* this can be achieved using a plate reader with shaking and heating capabilities, with OD and fluorescence readings taken automatically at regular time intervals (Hirst, 2014). This is more challenging for a thermophile as most plate readers do not have the ability to heat to the required temperature, and evaporation becomes more of a problem than it already is at 30 or 37 °C.

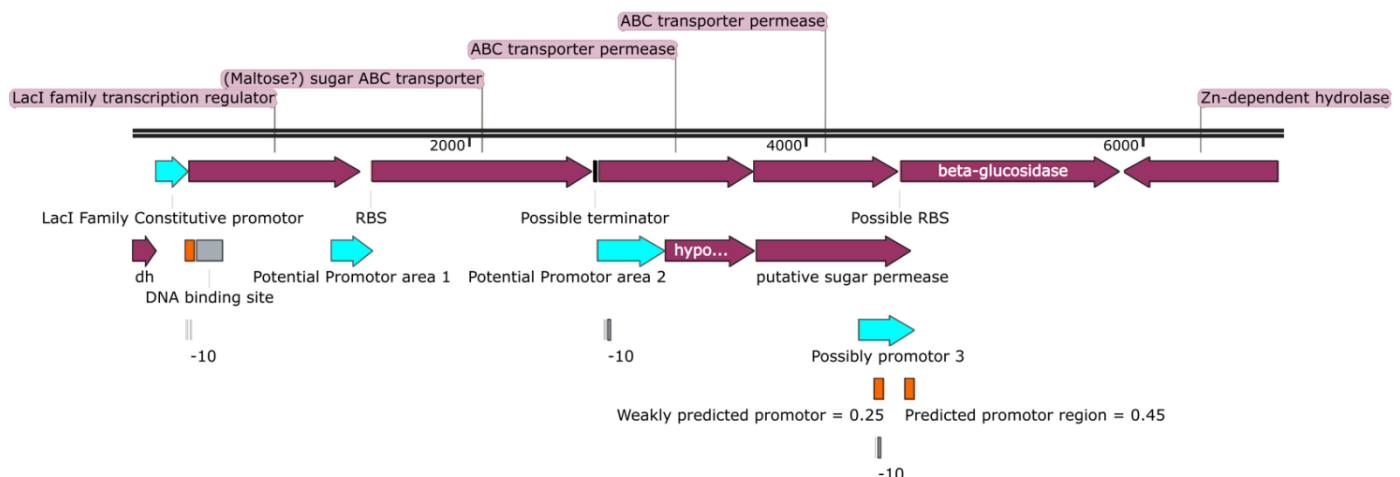
#### **6.4.3.2 Read-through from the PslpA-Kanamycin antibiotic resistance cassette.**

Promoter characterization using sfGFP shows weak expression with no promoter present, using the pBBTth vector (Figure 6-14 and Figure 6-15). This is likely a result of transcriptional read through from the PslpA promoter driving the kanamycin resistance cassette (Figure 6-9). While the BioBricks cloning site is insulated by two terminators, these are designed for use in *E. coli* and may be less functional in *T. thermophilus*. There is also a possible terminator region downstream of the *kan* gene. Future vector designs should take this into account. However the weak level of expression detected with no promoter present could prove useful where low levels of expression are desirable.

#### **6.4.3.3 An ideal promoter would be a strong inducible promoter.**

A strong inducible promoter would be a powerful tool for the use of *T. thermophilus* for whole-cell biocatalysis. This would allow the accumulation of biomass before induction, and the optimization of expression conditions. Unfortunately the currently available inducible promoters have a number of disadvantages making them unattractive for this purpose.

A system mirroring the T7 system used in *E. coli* would be attractive (Tabor, 2001), but requires a suitable thermostable viral RNA polymerase and a tightly controlled inducible promoter such as the *lac* promoter. It is tempting to look in the genome of *T. thermophilus* for systems resembling the *lac* operon (Jacob and Monod, 1961). On the pTT27 plasmid of *T. thermophilus*, a gene for a LacI family transcription regulator is followed by an operon ending in a gene for a beta-glucosidase (Figure 6-32). Possibly a promoter for this operon is regulated by the LacI family transcription regulator and may be inducible. The presence of non-coding regions and predicted promoter regions using bioinformatics tools identifies three potential areas which could contain such a promoter, which would be interesting to test using the sfGFP reporter system.



**Figure 6-32 – A *lac* like operon on the pTT27 plasmid of *T. thermophilus* HB27**

An area of the *T. thermophilus* HB27 pTT27 plasmid (bp 33,740 to 40,570 is shown), contains a region with a number of similarities to the *lac* operon. A gene for the expression of a LacI family transcription regulator is followed by an operon ending in the gene for a beta-glucosidase. It is unclear if the LacI family transcription regulator regulates a promoter for this operon, or indeed where this promoter is. However there are a number of likely options and regions that are identified as promoters using various bioinformatics tools available online (*Molbiol-Tools Web Portal*, 2016), or as they are non-coding regions. This figure was made with SnapGene.

#### **6.4.4 Using *T. thermophilus* for the hydrolysis of methyl *p*-toluate or *p*-toluamide by whole-cell biocatalysis.**

##### **6.4.4.1 Expression of AfEst2 in *T. thermophilus* for whole-cell biocatalysis**

Expression of AfEst2 in *T. thermophilus* was strong with the protein clearly visible by both SDS-PAGE and western blot analysis when expressed alone or as part of a synthetic operon (Figure 6-19 and Figure 6-21). In the synthetic operon the AfEst2 gene was approximately 3.5 kbp from the beginning of the mRNA transcript (Figure 6-16), yet a similar level of expression was detected compared to AfEst2 expressed alone. Translation initiation via the RBS in the middle of the operon mRNA transcript is equally as efficient as initiation from the same RBS near the beginning of the transcript for AfEst2 alone. This suggests the construction of a large synthetic operon for expression in *T. thermophilus* is a reasonable approach for the expression of multiple genes.

Slower growth by constructs designed for expression from PslpA is likely due to the burden of high expression on the *T. thermophilus* cells. Lower cell densities in these constructs also resulted in a poorer conversion of methyl *p*-toluate than expression from the weaker Prmp promoter. Induction of Pnar resulted in very slow growth rates and demonstrates this promoter to be unsuitable for the use of *T. thermophilus* for whole-cell biocatalysis.

*T. thermophilus* has a number of native esterase enzymes (Henne *et al.*, 2004), and it likely one or more of these have some activity against methyl *p*-toluate, as some *p*-toluic acid was detected in the control experiments.

##### **6.4.4.2 Expression of ss- $\gamma$ -lactamase in *T. thermophilus***

A preliminary test for the expression of ss- $\gamma$ -lactamase in *T. thermophilus* showed good expression of the enzyme (Figure 6-24), and that it could be used for the whole-cell conversion of *p*-toluamide to *p*-toluic acid, although yields were low (Figure 6-26). This further demonstrates the use of *T. thermophilus* for whole-cell biocatalysis although a more thorough characterization of the expression of ss- $\gamma$ -lactamase would be useful.

### **6.4.5 Aldehyde ferredoxin oxidoreductases limit the potential of *T. thermophilus* for the reduction of acids or aldehydes.**

#### **6.4.5.1 Aldehyde ferredoxin oxidoreductases**

The *T. thermophilus* genome contains three tungsten containing aldehyde ferredoxin oxidoreductases. These enzymes catalyze the reversible oxidation of aldehydes into carboxylic acids, generally using ferredoxin as a redox partner (Bever, 2008). They are part of a group of pyranopterin containing molybdo- or tungsto enzymes, in which tungsten is always part of the active site, held by two pyranopterin cofactors and an oxo group. These enzymes contain an iron-sulphur cluster which is responsible for the transfer of electrons from the pyranopterin cofactor to an electron acceptor, usually ferredoxin (Omelchenko *et al.*, 2005). While the reaction is reversible it has been shown to strongly favor the oxidative direction *in vitro* (Heider, Ma and Adams, 1995).

The AFOs in *T. thermophilus* have the accession numbers: GI:499487549, GI:499485830 and GI:499487802 (labelled as TTh Afo 1 to 3 respectively in Table 6-5). Some of the best studied AFOs are found in *Pyrococcus furiosus* which has five AFOs. These are labelled as: AOR - most active on aldehydes derived from amino acids, FOR – active against small C1-C3 aldehydes, GAPOR - only active against glyceraldehyde-3-phosphate, WOR4 – no known function and WOR5 - with high affinity for several substituted and non-substituted aliphatic and aromatic aldehydes with variable chain lengths (Bever, 2008). The percentage identity similarities of these AFOs compared to the *T. thermophilus* AFOs is compared in Table 6-5, determined using the Clustal Omega multiple sequence alignment tool (McWilliam *et al.*, 2013).

**Table 6-5 – Percentage identity matrix comparing the AFO's from *P. furiosus* with those of *T. thermophilus*.**

	WOR5	FOR	WOR4	AOR	GAPOR	TTh AFO 1	TTh AFO 2	TTh AFO 3
TTh AFO 1	20.96	27.51	30.97	35.71	34.52	100	39.39	34.3
TTh AFO 2	22.49	32.84	35.55	36.49	40.03	39.39	100	44.2
TTh AFO 3	23.93	30.11	29.95	33.97	38.15	34.3	44.2	100

AFO's from *P. furiosus* have been labelled as they are referred to in the literature as AOR, FOR, GAPOR, WOR4 and WOR5 (Bever, 2008). *T. thermophilus* AFO's are labelled as TTh Afo 1 to 3. Determined using the Clustal Omega multiple sequence alignment tool (McWilliam *et al.*, 2013).

#### 6.4.5.2 Expression of ApADH in *T. thermophilus* for the reduction of *p*-tolualdehyde or *p*-toluic acid.

*P. furiosus* has been used previously for the reduction of aromatic carboxylic acids through the action of its native AFOs, relying on an ADH to pull the equilibrium towards the alcohol product (Van Den Ban *et al.*, 1999). The redox state of the ferredoxin pool has been shown to be a critical factor in achieving reduction with AFOs, with higher levels of reduced ferredoxin favoring the reduction of acids (Van Den Ban *et al.*, 1999). Feeding *P. furiosus* hydrogen allowed hydrogenase enzymes to regenerate reduced ferredoxin improving this reaction (Ni *et al.*, 2013). It was hoped that the expression of ApADH might pull the equilibrium catalyzed by the AFOs through to *p*-tolyl alcohol, or at least compete with the AFOs allowing more *p*-tolualdehyde to be converted to *p*-tolyl alcohol rather than *p*-toluic acid. As the ferredoxin pool has been shown to be important for the AFO enzymes, and it is likely dependent on the growth phase of *T. thermophilus*, efforts were made to synchronize the growth of *T. thermophilus* cultures expressing ApADH and an empty plasmid control.

*p*-toluic acid and *p*-tolualdehyde were added in the early exponential phase in the hope growing cultures would be producing more reduced ferredoxin (Figure 6-25). Negligible conversion of *p*-toluic acid was seen suggesting that the AFOs had very

little reductive activity. The expression of ApADH in *T. thermophilus* had little impact on its conversion of *p*-tolualdehyde, with native ADH enzymes responsible for the most of conversion to *p*-tolyl alcohol observed (however the majority was still oxidized to *p*-toluic acid). ApADH has an optimum temperature > 90 °C so may not be as active at 70 °C as native *T. thermophilus* ADH enzymes (Hirakawa, Kamiya and Kawarabayashi, 2004). Marginally more *p*-tolualdehyde was converted to *p*-tolyl alcohol by the *T. thermophilus* expressing ApADH from the PslpA promoter, possibly as a result of the expression of this enzyme.

#### **6.4.5.3 Could the reduced ferredoxin pool be increased in *T. thermophilus*?**

For *P. furiosus* to reduce carboxylic acids, maintaining high levels of reduced ferredoxin was shown to be important (Ni *et al.*, 2013). If high levels of reduced ferredoxin could be maintained in *T. thermophilus*, possibly the AFOs could be used in a similar way, in the reductive direction rather than the oxidative. A genome scale flux balance model of *T. thermophilus* has been developed and is available (Lee *et al.*, 2014). This may offer potential routes for metabolic engineering to try and increase the reduced ferredoxin pool.

A NADH-ferredoxin oxidoreductase from *Trichomonas vaginalis* has previously been expressed in *E. coli* for production of reduced ferredoxin using NADH as an electron donor (Do *et al.*, 2009). Possibly a similar enzyme might be used in *T. thermophilus* to produce more reduced ferredoxin.

#### **6.4.6 Gene deletion techniques for *T. thermophilus***

##### **6.4.6.1 Deletion of *afo* genes may be required for the pathway to function in *T. thermophilus*.**

For effective incorporation of any reductive biotransformation using *T. thermophilus* and featuring aldehydes in the pathway, the deletion of the three *afo* genes may be essential to stop oxidation of the aldehyde into its corresponding carboxylic acid. Cas9 may offer an interesting alternative for gene deletion to the more conventional use of the homologous recombination with *pyrE* gene used as a marker.

#### **6.4.6.2 Cas9 could offer an efficient method for gene deletion in *T. thermophilus*.**

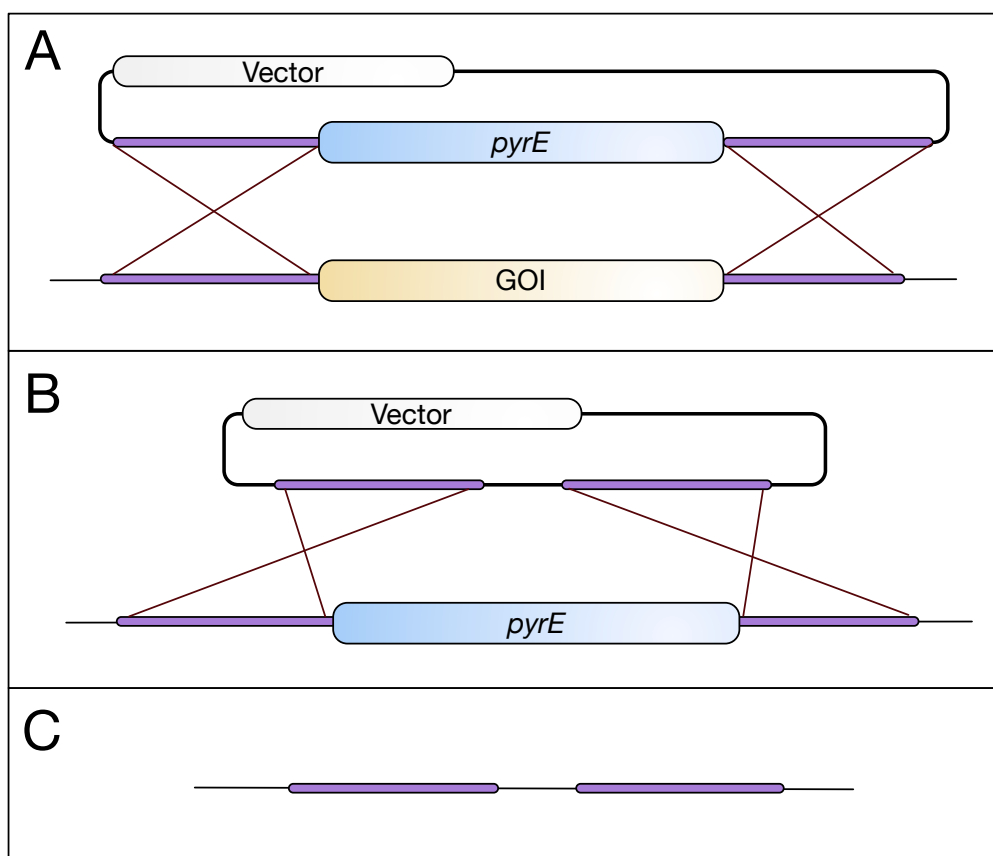
Recently CRISPR-Cas9 has received much attention as an efficient method for gene deletion, among other things (Mali, Esvelt and Church, 2013). Possibly the CRISPR-Cas9 system may be used for the deletion of the *afo* genes from *T. thermophilus*. In order to begin exploring this possibility we have looked at expressing a well characterized Cas9 from *Streptococcus thermophilus* (Xu *et al.*, 2015). Interestingly, when constructs were created for the expression of Cas9 from strong promoters PslpA and Prmp, either no colonies, or only a couple of colonies, were recovered following transformation into *T. thermophilus*. However many colonies were recovered when a construct for the expression of Cas9 from the much weaker Pnqo promoter was transformed into *T. thermophilus*. It seems likely that high levels of Cas9 expression is toxic to *T. thermophilus*, possibly due to non-specific cutting of DNA by the Cas9 protein (Hsu, Lander and Zhang, 2014). The expression Cas9 was detected by western blot analysis when expressed from Pnqo (Figure 6-30). The next steps will be the design of an RNA to allow Cas9 targeting, and validation of the Cas9 gene deletion approach. A good initial target may be the *pyrE* gene which would result in uracil autotrophy but resistance to 5-fluoroorotic acid (5-FOA) (Tamakoshi *et al.*, 1997), facilitating screening for a successful gene deletion.

#### **6.4.6.3 PyrE gene deletion technique**

An approach known to work with *T. thermophilus* is homologous recombination using the *pyrE* gene as a marker, shown in Figure 6-33. A *pyrE* knockout is first prepared. Regions of homology to the target gene, 500 bp in size, are prepared either side of the *pyrE* gene on a vector, with homologous recombination replacing the target gene with the *pyrE* gene in the chromosome. The *pyrE* gene is then deleted again to allow subsequent rounds of gene deletion.

This technique makes use of the uracil autotrophy and resistance 5-FOA that occurs upon *pyrE* gene deletion, with growth in the absence of uracil restored with the knock-in of *pyrE* (Tamakoshi *et al.*, 1999).





**Figure 6-33 – The *pyrE* gene deletion method.**

500 bp regions of homology to a target gene (GOI) are prepared flanking a *pyrE* gene on a vector. Homologous recombination restores growth in the absence of uracil of a *pyrE* knockout strain of *T. thermophilus*, with the target gene (GOI) replaced with the *pyrE* gene. The *pyrE* gene is deleted by homologous recombination allowing the next round of gene deletion making use of *pyrE*. Target gene (GOI) is deleted and the *pyrE* knockout is maintained.

### 6.4.7 Conclusions

The ease with which *T. thermophilus* can be transformed has driven its use as a model organism for understanding life at high temperatures (Cava, Hidalgo and Berenguer, 2009), and could also drive its uptake as the host organism of choice for whole-cell biocatalysis at elevated temperatures. Having a number of genetic parts that can be combined predictably is an important step in developing *T. thermophilus* as a host organism.

Better characterization of a wider range of promoters should be performed, in order to better understand the regulation on the current constitutive promoters, and in an

effort to find an industrially useful inducible promoter. The RBS calculator allows strong RBSs to be designed for each CDS. Codon optimization is essential for the expression of genes in *T. thermophilus*. However more work is needed to understand translation efficiency determinants, although this is not a problem unique to *T. thermophilus*. The approach taken by DNA 2.0 for *E. coli* and other organisms may offer a better approach for designing genes for *T. thermophilus* (Welch *et al.*, 2009).

Whole-cell biocatalysis by *T. thermophilus* cells expressing AfEst2 and ss- $\gamma$ -lactamase has been demonstrated. However the completion of the entire pathway shown in Figure 6-5 would offer a much better demonstration. The three *afo* genes in *T. thermophilus* can be deleted to allow this pathway to be completed, either using a Cas9 approach or using the *pyrE* gene deletion method (Tamakoshi *et al.*, 1999) (Figure 6-33).

This study highlights the potential of *T. thermophilus* as an organism for synthetic biology and whole-cell biocatalysis at elevated temperatures. The development of well characterized parts and tools, and the completion of the pathway in Figure 6-5 would help that potential be realized.

## **Chapter 7 - General Discussion**

## 7.1 Multistep pathways in biocatalysis

This study has sought to characterise a number of enzymes before incorporating them into a multistep enzyme reaction. This multistep reaction has then been carried out *in vitro* using isolated enzymes, and *in vivo* exploring the use of the thermophile *T. thermophilus* as a host organism.

Currently single step enzyme reactions are most commonly used for industrial biocatalysis. However as the uptake of enzymes for synthetic chemistry continues, multistep reactions are likely to become more prominent (Bruggink, Schoevaart and Kieboom, 2003; Wohlgemuth, 2010). The complexity of these reactions increases substantially as more steps are added, requiring more of a synthetic biology approach to their design and optimization than a single step reaction might require (Ricca, Brucher and Schrittwieser, 2011).

Synthetic biology is inspired by engineering disciplines demanding the use of well characterized “parts” (Ellis, Adie and Baldwin, 2011). In a similar vein, the enzyme toolbox is an image often invoked to describe the enzymes which have established themselves as tools for synthetic chemistry (Meyer *et al.*, 2013). Possibly the enzymes in the enzyme toolbox could be thought of as “parts” for combination into predictable multistep cascade reactions, although the thorough biochemical characterization of each enzyme would first be necessary. The characterisation of enzymes for their incorporation into a multistep reaction is the approach taken throughout this study. A lesson that has been learnt in synthetic biology is that each “part” must be robustly characterized before it is used in a larger “device” (Cheng and Lu, 2012). In the same way, each enzyme in a toolbox needs thorough characterization before it can be incorporated into a larger pathway.

## 7.2 Characterising enzymes

Possibly the most important characterisation of enzymes is that of their substrate specificity. However, before assays examining substrate specificity can be carried out, constraints on the assay conditions such as pH and temperature must be determined. Furthermore, some substrates may only be accepted at higher temperatures or different pH values, with some enzymes catalysing different reactions entirely at extremes of pH (Höllrigl *et al.*, 2008).

Often substrate specificity is reported as the relative activity at a single substrate concentration, or the percentage conversion in a specified time period. Arguably this provides little more than a yes/no analysis of whether a substrate is turned over by an enzyme. Far more meaningful information can be gained looking at the kinetics of the enzyme with changing substrate concentration. This is particularly important for modelling complex reactions, as some substrates may only have industrially relevant rates at high concentrations.

Testing all available substrates for an enzyme is unfeasible, so some predictions on substrate specificity must be made. Given enough data, quantitative structure activity relationship (QSAR) models might be constructed to help extrapolate substrate specificity (Damborsky and Brezovsky, 2014). Alternatively, some progress is being made on predicting substrate specificity from the crystal structures of some enzymes, although protein flexibility must be taken into account (Tyagi and Pleiss, 2006).

Characterisation of product inhibition is also important, as we showed for the CAR enzymes. Enzymes should be tested for product inhibition, as well as the effects of other compounds likely to be included in a multistep reaction such as cofactors or metal ions.

Other important considerations when characterizing enzymes are the effects of pH and temperature. Characterizing the effect of pH on the activity of enzymes is often reported at a single substrate concentration, and the enzymes in this study were characterized in this way. However, the effect of pH on enzyme activity is more complex than this, as it is a result of the changing protonation of various amino acid residues in the protein, or even the protonation of substrate itself. It must be borne in mind that accurately reporting these effects requires a prohibitively large number of experiments to determine Michaelis-Menten parameters at each pH. Furthermore, substrates may be affected in different ways at varying pH values (Stryer, 1995). These issues mean that characterization of pH effects on enzymes must necessarily be a compromise of data range and practicality. The effects of temperature on enzyme activity is also often reported at a single substrate concentration. However, unless changing temperature causes a conformational change in the enzyme,

activity will generally follow the Arrhenius equation which can be incorporated into Michaelis-Menten-like equations in the place of  $k_{cat}$  (Laidler, 1984).

Thermostability can be reported in a number of ways, such as looking at residual activity after heating the enzyme at a range of temperatures, or looking for the melting temperature using techniques such as dynamic scanning fluorimetry (DSF) (Vivoli *et al.*, 2014). In practice, enzymes are generally used at temperatures well below these limits. What is often more relevant is operational stability, which temperature will often play a large role in. Operational stability is generally reported as the half-life of the enzyme in a certain condition. This might be considered as a simplified view of enzyme degradation, which is itself a complex process (Polizzi *et al.*, 2007).

Possibly the creation of enzyme specification sheets, similar to specification sheets produced on DNA parts for synthetic biology, might aid the design of novel multistep pathways (Cooling *et al.*, 2010; Sainz de Murieta, Bultelle and Kitney, 2016).

Such a sheet might seek to contain information on changing activity with pH and temperature, any inhibition by commonly used buffers, metal ions or cofactors and information on stability at extremes of temperature, pH or in solvents, as well as operational stability in more commonly used conditions. Where structural information on an enzyme is available, limitations on the size and shapes of molecules which might fit in the active site could be provided. Providing information on substrate specificity in terms of a structure-activity relationship would also be valuable in allowing chemists to select suitable enzymes for specific reactions with greater confidence.

### **7.3 The characterisation of the thermostable esterase Af-Est2 for inclusion in the multistep reaction.**

The first step in the multistep reaction shown in Figure 1-8 was filled with a novel esterase, Af-Est2, from the hyperthermophile *A. fulgidus*. While hydrolase enzymes are the most studied of the six main enzymes classes, thermostable hydrolase enzymes offer unparalleled stability with novel thermostable hydrolase enzymes still sought after industrially (Littlechild, 2015). In this respect, Af-Est2 showed not only

excellent thermostability but maintained high levels of activity in extremes of pH and at high solvent concentrations, making this enzyme very industrially attractive. The excellent solvent stability of Af-Est2 is particularly intriguing as this may facilitate the use of Af-Est2 in the absence of water for bond forming condensation reactions (Klibanov, 2001).

Characterization of the substrate specificity of Af-Est2 is limited only to methyl *p*-toluate and the standard *p*-nitrophenyl esters which are often used to examine the substrate specificity of esterase or lipase enzymes. A more thorough characterization of substrate specificity could be beneficial in characterizing this enzyme for industrial use. The crystal structure of Af-Est2 might be useful in selecting a panel of substrates for this purpose, and possibly in predicting other substrate specificities. Furthermore, the crystal structure, biochemical characterization and high stability of Af-Est2 provides an excellent starting point for engineering of the enzyme for industrial use, such as expanding substrate specificity (Dalby, 2007).

## **7.4 Characterisation of carboxylic acid reductase enzymes.**

The second step in the multistep reaction was filled by a CAR enzyme (Figure 1-8). Previous work on CARs has mostly focused on the CAR from *Nocardia iowensis* (He *et al.*, 2004), and to a lesser extent the CAR from *Mycobacterium marinum* (Akhtar, Turner and Jones, 2013). Recently, some characterization was carried out on two other CARs, but this was limited to a yes/no screen of only a few substrates, with products confirmed by GC/MS analysis (Moura *et al.*, 2015). Since only a few CARs have been explored, with often a limited biochemical characterization, we sought to characterize a number of CARs from across the CAR family.

Our work identified a large number of CAR sequences by bioinformatics, with a careful phylogenetic analysis carried out to examine the relationship between these sequences. CARs from across the family were selected for characterization. Kinetic characterization of these CARs strongly suggested that the first step in the proposed reaction scheme is crucial for substrate specificity, and that while substrate specificity is fairly broad, it is similar across all the CARs. We also showed ordered

binding of substrates at the adenylation domain, similar to other members of the ANL superfamily which we identified CARs to be a member of (Gulick, 2009).

CARs offer the reduction of carboxylic acids in mild conditions in stark contrast to chemical methods typically requiring harsh reducing agents in stoichiometric quantities. Moreover they offer selective reduction to only the aldehyde product rather than complete reduction to the alcohol product as is generally the case using chemical methods (Napora-Wijata, Strohmeier and Winkler, 2014). For these reasons CARs are promising enzymes and could be thought of as one the next enzymes to achieve a place in the enzyme toolbox.

For example, CARs are already being used in the construction of numerous multistep enzymes pathways, both for biotransformations and for fermentations (Hansen *et al.*, 2009; Kallio *et al.*, 2014; Napora-Wijata *et al.*, 2014). Furthermore CARs join up a number of established enzymes in the enzyme toolbox, and are a crucial link in constructing a network of biocatalysis reactions, important for the construction of industrial multistep reactions.

However, while CARs are very attractive enzymes for synthetic chemistry, other enzyme classes in the enzyme toolbox feature a number of distinct enzymes able to meet a wide range of requirements (Turner and O'Reilly, 2013). The CARs we tested cover most of the known family of CARs, all from the Actinomyces. The similarities of substrate specificity and operating conditions between these enzymes could be a limitation in the CARs ability to meet the wide range of requirements the other enzymes classes in the toolbox are capable of. Future efforts to engineer CARs are likely to expand the substrate specificity and operating conditions. Additionally, a CAR from the fungus *Syncephalastrum racemosum* was recently identified (Brenna *et al.*, 2015), and possibly more CARs from other sources will follow, resulting in more diverse set of CAR enzymes.

## 7.5 Completion of a multistep enzymatic reaction

Af-Est and mpCAR, the most naturally thermostable CAR identified, were used to make up the first two steps of the three enzyme pathway outlined in Figure 1-8. An ADH from the hyperthermophile *A. pernix*, apADH, has been characterized previously and filled the final step in this pathway (Guy, Isupov and Littlechild, 2003;



Hirakawa, Kamiya and Kawarabayashi, 2004). With a pathway of well characterized enzymes, the *in vitro* and *in vivo* reaction were both investigated.

## **7.6 The use of isolated enzymes with mathematical modelling and cofactor regeneration**

One of the ambitions of synthetic biology is the *in silico* design of organisms. This ambition might be shared with the field of biocatalysis for the design of multistep pathways for synthetic chemistry. This is probably most achievable in the design of reactions featuring isolated enzymes, as the unpredictable components of whole-cells are removed.

The optimization of multistep reactions is challenging and will generally require some form of modelling. This can be either a data driven approach such as the generation of a response surface, or mechanistic modelling using parameters determined experimentally to fit equations such as the Michaelis-Menten equation (Santacoloma *et al.*, 2011). The use of mechanistic modelling also facilitates understanding of the reaction being modelled, as we demonstrated with the identification product inhibition in the CAR step and its removal with the addition of an inorganic pyrophosphatase enzyme. Mechanistic models might also be reused in other reactions, something that is not possible using data driven modelling. A further three enzymes were included for the regeneration of NADPH, NADH and ATP. This included the identification and characterization of a novel thermostable AMP polyphosphate phosphotransferase enzyme (PAP). However, characterization of these enzymes was fairly limited, with the scope for a much more thorough characterization.

Modelling of the ATP regeneration system posed the most challenges and could not be achieved entirely faithfully. It would be interesting if the modelling of free and bound magnesium ions could be achieved, and how this might affect the model predictions. A more efficient ATP regeneration system might seek to mimic the way ATP is more commonly regenerated in nature, through the use of an ATP synthase embedded in a synthetic membrane separating a pH gradient (Nam, Struck and Holtzapple, 1996).

This work used “good modelling practice” which included uncertainty and sensitivity analysis (Sin, Gernaey and Lantz, 2009), often omitted in other similar work. This allowed the impact of assumptions and unknowns to be evaluated and taken into account. Using uncertainty and sensitivity analysis offers a very useful approach to building models quickly from limited data. Even if a mechanistic model will ultimately not be used, with appropriate bounds for unknown parameters they can help identify which experiments should be carried out. For example this approach might be used to inform which variables should be tested using a response surface methodology.

Where a mechanistic model has been developed for a process, it can be exploited to optimise the reaction towards a desired goal. Numerous methods for global optimisation are available, with genetic algorithms offering a powerful tool for avoiding local minima or maxima, and for multi-objective problems likely to arise for biocatalysis. In this study the model constructed was used in this way to minimize enzyme concentrations while reaching a target product concentration in a specified time, demonstrating the power of this approach.

Whether for *in vitro* or *in vivo* biotransformations, modelling offers a powerful tool for the design and optimisation of reactions. For *in vivo* models, ideally enzyme characterisation could be carried out in a buffer mimicking the cellular environment. However the use of *in vivo* models to realise optimised reactions is limited by our inability to accurately predict protein production rates, mostly at the level of transcription.

Reaching an optimised process is an important goal for biocatalysis, but where this is the only goal and the process is relatively well understood, it is likely faster and simpler to use a data driven approach such as response surface methodology. However the use of a mechanistic model offers a number of advantages if a process is under active development, as it allows options to be explored quickly *in silico*. Both methods have their place for process modelling of biocatalysis reactions.

## **7.7 Whole-cell biocatalysis at elevated temperatures.**

*T. thermophilus* is an attractive host organism for whole-cell biocatalysis at elevated temperatures primarily due to its natural competence (Cava, Hidalgo and Berenguer, 2009), making transformations much easier to achieve than in other thermophilic

organisms. The main competitor for this role is likely *Geobacillus stearothermophilus* (or other *Geobacillus* species), which is being actively pursued as a thermophilic host organism for various synthetic biology applications (French, 2009). An ability to grow at a wide range of growth temperatures, utilizing a range of carbon sources and being a relative to the well-studied mesophilic bacteria *Bacillus subtilis* make this an attractive host organism (Takami *et al.*, 2004; Olson *et al.*, 2012; Kananavičiute and Čitavičius, 2015). However in many cases development has been slowed substantially by difficulties in transforming the organism, often requiring complex protocols including electroporation (Kananavičiute and Čitavičius, 2015).

For both *T. thermophilus* and *G. stearothermophilus* the development of synthetic biology tools and parts is critical to their success (Kuhn *et al.*, 2010). In this work the available genetic tools and parts for *T. thermophilus* were reviewed. sfGFP expression allowed three constitutive promoters to be easily characterized. A robust protocol should be set up to allow the characterization of more promoters to cover a range of expression strengths, with the development of synthetic promoters another possibility. Ideally a strong but inducible promoter would allow the accumulation of biomass followed by high levels of protein expression, with some potential promoters identified in this study.

While a good start was made on the development of synthetic biology tools for *T. thermophilus*, a demonstration of its use as a host for whole-cell biocatalysis was more limited. Due to the presence of native AFO enzymes for the oxidation of aromatic aldehydes in *T. thermophilus*, the transformation was limited to only hydrolysis of methyl *p*-toluate. Potentially the native AFO enzymes might be made to work in the reverse direction, but it is more likely that their deletion will be necessary for the complete transformation to be successful. Exploring the use of Cas9 in *T. thermophilus* would be an interesting approach for the deletion of these genes.

Possibly a more achievable short term goal might be the expression of the three enzymes making up the multistep reaction in *E. coli*. An interesting proposal might be the use of the mathematical model developed for the *in vitro* reaction, in informing the design of a new synthetic operon for expression in *E. coli*.

## 7.8 Biocatalysis and synthetic biology in the future

Biology is said to be going through a reproducibility crisis, but is also changing at a fast pace. One positive change is that some journals are now asking that all data associated with publications be uploaded along with the manuscript. Automation is becoming more accessible, both in-house automation and cloud labs such as Transcriptic, offering more reproducible experiments. Programming languages for biology such as Antha are being developed, aimed at describing protocols in an exact way and integrating with liquid handling robots. Companies such as Biobright are producing hardware and software tools aimed to change the way biology is done in the lab, collecting data from new sensors or lab equipment in an integrated way.

For both biocatalysis and synthetic biology, high-throughput experiments are becoming more routine. The falling price of synthetic DNA means it's easy for large numbers of genes, or even synthetic sequences such as ancCAR, to be tested quickly and cheaply, especially with increasing levels of lab automation.

It has been remarked that biocatalysis often constitutes the second generation choice for synthetic chemistry processes, generally due to slower development times. The use of characterised parts could take out some of the trial and error present in the approach of screening large numbers of enzymes to fit a reaction. Furthermore as the enzyme toolbox grows, possibly there will be less need to engineer enzymes to fit specific roles, as a suitable enzyme may already be present.

The design of multistep pathways could be aided greatly using computer tools armed with knowledge of the enzyme reactions available, and their strengths and limitations. Such a tool would be useful in suggesting to chemists where biocatalysis might offer a powerful solution to a synthesis.

## 7.9 Conclusions

With the increasing uptake of enzymes for synthetic chemistry, and a number of advantages to be gained from combining multiple enzyme reactions into a single process, multistep reactions are likely to become more common in the future. However multistep reactions are generally more complex than single step reactions with many variables to consider. Thorough characterization of enzymes such as has

been performed in this study will be critical in the future design of novel multistep enzyme reactions. Mathematical modelling offers a powerful tool for the optimization and understanding of these reactions, which the work in this study demonstrates well with the use of good modelling practice. Finally where there is a need to use enzymes at elevated temperatures, the use of a well-developed thermophilic host organism for whole-cell biocatalysis such as *T. thermophilus* will allow these types of reactions to be considered. The work in this study is a strong start in developing *T. thermophilus* as a host organism.



## **Chapter 8 - Bibliography**

- Abascal, F., Zardoya, R. and Posada, D. (2005) 'ProtTest: Selection of best-fit models of protein evolution', *Bioinformatics*, 21(9), pp. 2104–2105.
- Agarwal, P. K., Webb, S. P. and Hammes-Schiffer, S. (2000) 'Computational Studies of the Mechanism for Proton and Hydride Transfer in Liver Alcohol Dehydrogenase', (4), pp. 4803–4812.
- Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M. and Pearl, L. H. (1997) 'Crystal structure of the beta-glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability.', *Journal of molecular biology*, 271(5), pp. 789–802.
- Akhtar, M. K., Turner, N. J. and Jones, P. R. (2013) 'Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities.', *Proceedings of the National Academy of Sciences of the United States of America*, 110(1), pp. 87–92.
- Aldridge, S. (2013) 'Industry backs biocatalysis for greener manufacturing', *Nature Biotechnology*, 31(2), pp. 95–96.
- Allouche, N., Fki, I. and Sayadi, S. (2004) 'Toward a High Yield Recovery of Antioxidants and Purified Hydroxytyrosol from Olive Mill Wastewaters', *Journal of Agricultural and Food Chemistry*, 52(2), pp. 267–273.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool.', *Journal of molecular biology*, 215(3), pp. 403–410.
- Anastas, P. T. and Warner, J. C. (1998) *Green Chemistry: Theory and Practice*, Oxford University press. Oxford university press.
- Angkawidjaja, C., Koga, Y., Takano, K. and Kanaya, S. (2012) 'Structure and stability of a thermostable carboxylesterase from the thermoacidophilic archaeon *Sulfolobus tokodaii*', *FEBS Journal*, 279(17), pp. 3071–3084.
- Aoki, K. and Itoh, T. (2007) 'Characterization of the ColE2-like replicon of plasmid pTT8 from *Thermus thermophilus*', *Biochemical and Biophysical Research Communications*, 353(4), pp. 1028–1033.
- Armstrong, J. M., Myers, D. V., Verpoorte, J. A. and Edsall, J. T. (1966) 'Purification and properties of human erythrocyte carbonic anhydrases.', *The Journal of biological chemistry*, 241(21), pp. 5137–5149.
- Atomi, H., Imanaka, T. and Fukui, T. (2012) 'Overview of the genetic tools in the Archaea', *Frontiers in Microbiology*, 3, pp. 1–13.
- Atomi, H., Sato, T. and Kanai, T. (2011) 'Application of hyperthermophiles and their enzymes', *Current Opinion in Biotechnology*, 22(5), pp. 618–626.
- Auerbach, G., Huber, R., Grättinger, M., Zaiss, K., Schurig, H., Jaenicke, R. and Jacob, U. (1997) 'Closed structure of phosphoglycerate kinase from *Thermotoga maritima* reveals the catalytic mechanism and determinants of thermal stability.', *Structure*, 5(11), pp. 1475–1483.
- Averhoff, B. (2004) 'DNA Transport and Natural Transformation in Mesophilic and Thermophilic Bacteria', *Journal of Bioenergetics and Biomembranes*, 36(1), pp. 25–33.
- Averhoff, B. (2006) 'Genetic Systems for *Thermus*', in *Methods in Microbiology*. Academic Press, pp. 279–308.
- Averhoff, B. (2009) 'Shuffling genes around in hot environments: The unique DNA transporter of *Thermus thermophilus*', *FEMS Microbiology Reviews*, 33(3), pp. 611–626.
- Bains, J., Kaufman, L., Farnell, B. and Boulanger, M. J. (2011) 'A product analog bound form of 3-oxoadipate-enol-lactonase (PcaD) reveals a multifunctional role for the divergent cap domain', *Journal of Molecular Biology*, 406(5), pp. 649–658.



- Van Den Ban, E. C. D., Willemsen, H. M., Wassink, H., Laane, C. and Haaker, H. (1999) 'Bioreduction of carboxylic acids by *Pyrococcus furiosus* in batch cultures.', *Enzyme and Microbial Technology*, 25, pp. 251–257.
- Banner, T., Fosmer, A., Jessen, H., Marasco, E., Rush, B., Veldhouse, J., De Souza, M., Tao, J. and Kazlauskas, R. (2011) *Biocatalysis for Green Chemistry and Chemical Process Development*. Edited by J. A. Tao and R. Kazlauskas. John Wiley and Sons.
- Basen, M., Sun, J. and Adams, M. W. W. (2012) 'Engineering a Hyperthermophilic Archaeon for Temperature Dependent Product Formation', *mBio*, 3(2), pp. 1–8.
- Bencharit, S., Edwards, C. C., Morton, C. L., Howard-Williams, E. L., Kuhn, P., Potter, P. M. and Redinbo, M. R. (2006) 'Multisite Promiscuity in the Processing of Endogenous Substrates by Human Carboxylesterase 1', *Journal of Molecular Biology*, 363(1), pp. 201–214.
- Benner, S. A., Yang, Z. and Chen, F. (2011) 'Synthetic biology, tinkering biology, and artificial biology. What are we learning?', *Comptes Rendus Chimie*, 14(4), pp. 372–387.
- Berg, J. M., Tymoczko, J. L. and Stryer, L. (2010) *Biochemistry*. 7th edn. New York: W.H. Freeman & Co Ltd.
- Berkner, S. and Lipps, G. (2008) 'Genetic tools for *Sulfolobus* spp.: Vectors and first applications', *Archives of Microbiology*, 190(3), pp. 217–230.
- Berríos-Rivera, S. J., Bennett, G. N. and San, K.-Y. (2002) 'Metabolic Engineering of *Escherichia coli*: Increase of NADH Availability by Overexpressing an NAD<sup>+</sup>-Dependent Formate Dehydrogenase', *Metabolic Engineering*, 4(3), pp. 217–229.
- Beyers, L. E. (2008) *Tungsten Biochemistry in *Pyrococcus furiosus**. Technische Universiteit Delft.
- Bhaskar Khant, J. V. and Periasamy, M. (1991) 'Selective Reduction of Carboxylic Acids into Alcohols Using NaBH<sub>4</sub> and I<sub>2</sub>', *Journal of Organic Chemistry*, 56(13), pp. 5964–5965.
- Bhosale, S. H., Rao, M. B. and Deshpande, V. V. (1996) 'Molecular and industrial aspects of glucose isomerase.', *Microbiological reviews*, 60(2), pp. 280–300.
- Billig, S., Oeser, T., Birkemeyer, C. and Zimmermann, W. (2010) 'Hydrolysis of cyclic poly(ethylene terephthalate) trimers by a carboxylesterase from *Thermobifida fusca* KW3', *Applied Microbiology and Biotechnology*, 87(5), pp. 1753–1764.
- Blair, J. M. (1970) 'Magnesium, potassium, and the adenylate kinase equilibrium. Magnesium as a feedback signal from the adenine nucleotide pool', *European Journal of Biochemistry*, 13(2), pp. 384–390.
- Blow, D., Birktoft, J. and Hartley, B. (1969) 'Role of a buried acid group in the mechanism of action of chymotrypsin.', *Nature*, pp. 337–40.
- De Bont, J. A. m (1998) 'Solvent-tolerant bacteria in biocatalysis', *Trends in Biotechnology*, 16(12), pp. 493–499.
- Bonting, C. F. C., Kortstee, G. J. J. and Zehnder, A. J. B. (1991) 'Properties of polyphosphate - AMP phosphotransferase of *Acinetobacter* strain-210A', *Journal of Bacteriology*, 173(20), pp. 6484–6488.
- Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C. and Robins, K. (2012) 'Engineering the third wave of biocatalysis.', *Nature*, 485(7397), pp. 185–94.
- Bourne, P. C., Isupov, M. N. and Littlechild, J. A. (2000) 'The atomic-resolution structure of a novel bacterial esterase', *Structure*, 8(2), pp. 143–151.
- Brenna, E., Cannavale, F., Crotti, M., Parmeggiani, F., Romagnolo, A., Spina, F. and Varese, G. C. (2015) 'Biocatalysed reduction of carboxylic acids to primary alcohols in aqueous medium: A novel synthetic capability of the zygomycete fungus *Syncephalastrum racemosum*', *Journal of Molecular Catalysis B: Enzymatic*, 116, pp. 83–88.

- Brouns, S. J. J., Wu, H., Akerboom, J., Turnbull, A. P., De Vos, W. M. and Van Der Oost, J. (2005) 'Engineering a selectable marker for hyperthermophiles', *Journal of Biological Chemistry*, 280(12), pp. 11422–11431.
- Brown, W. G. and Chaikin, S. W. (1949) 'Reduction of Aldehydes, Ketones and Acid Chlorides by Sodium Borohydride', *Journal of the American Chemical Society*, 71(1), pp. 122–125.
- Bruggink, A., Roos, E. C. and de Vroom, E. (1998) 'Penicillin acylase in the industrial production of beta-lactam antibiotics', *Organic Process Research and Development*, 2(2), pp. 128–133.
- Bruggink, A., Schoevaart, R. and Kieboom, T. (2003) 'Concepts of nature in organic synthesis: Cascade catalysis and multistep conversions in concert', *Organic Process Research and Development*, 7(5), pp. 622–640.
- Buchan, J. R. and Stansfield, I. (2007) 'Halting a cellular production line: responses to ribosomal pausing during translation.', *Biology of the cell / under the auspices of the European Cell Biology Organization*, 99(9), pp. 475–487.
- Bujara, M., Billerbeck, S., Greve, F. and Panke, S. (2010) 'Synthetic Biology for Biocatalysis', in Timmis, K. N. (ed.) *Handbook of Hydrocarbon and Lipid Microbiology*. Berlin, Heidelberg: Springer, pp. 2939–2950.
- Bujara, Schümperli, M., Billerbeck, S., Heinemann, M. and Panke, S. (2010) 'Exploiting cell-free systems: Implementation and debugging of a system of biotransformations', *Biotechnology and Bioengineering*, 106(3), pp. 376–389.
- Cannarrozzi, G., Schraudolph, N. N., Faty, M., von Rohr, P., Friberg, M. T., Roth, A. C., Gonnet, P., Gonnet, G. and Barral, Y. (2010) 'A role for codon order in translation dynamics', *Cell*, 141(2), pp. 355–367.
- Cao, H., Han, H., Li, G., Yang, J., Zhang, L., Yang, Y., Fang, X. and Li, Q. (2012) 'Biocatalytic synthesis of poly( $\delta$ -Valerolactone) using a thermophilic esterase from *archaeoglobus fulgidus* as catalyst', *International Journal of Molecular Sciences*, 13(10), pp. 12232–12241.
- Carlson, R. (2016) 'Estimating the biotech sector's contribution to the US economy', *Nature Publishing Group*, 34(3), pp. 247–255.
- De Carvalho, C. C. C. (2011) 'Enzymatic and whole cell catalysis: Finding new strategies for old processes', *Biotechnology Advances*, 29(1), pp. 75–83.
- Cau, L. (2006) *Carrier-bound Immobilised Enzymes: Principles, Applications and Design*. Weinheim: Wiley-VCH.
- Cava, F., Hidalgo, A. and Berenguer, J. (2009) 'Thermus thermophilus as biological model', *Extremophiles*, 13, pp. 213–231.
- Cava, F., Laptenko, O., Borukhov, S., Chahlafl, Z., Blas-Galindo, E., Gómez-puertas, P. and Berenguer, J. (2007) 'Control of the respiratory metabolism of *Thermus thermophilus* by the nitrate respiration conjugative element NCE', *Molecular Microbiology*, 64(3), pp. 630–646.
- Cava, F., De Pedro, M. A., Blas-Galindo, E., Waldo, G. S., Westblade, L. F. and Berenguer, J. (2008) 'Expression and use of superfolder green fluorescent protein at high temperatures in vivo: A tool to study extreme thermophile biology', *Environmental Microbiology*, 10(3), pp. 605–613.
- Cava, F., Zafra, O., Magalon, A., Blasco, F. and Berenguer, J. (2004) 'A new type of NADH dehydrogenase specific for nitrate respiration in the extreme thermophile *Thermus thermophilus*', *Journal of Biological Chemistry*, 279(44), pp. 45369–45378.
- Chan, D. I. I. and Vogel, H. J. J. (2010) 'Current understanding of fatty acid biosynthesis and the acyl carrier protein', *Biochemical Journal*, 430(1), pp. 1–19.
- Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. and Rees, D. C. (1995) 'Structure of a

hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase.', *Science*, 267(5203), pp. 1463–1469.

Charpentier, J. (2002) 'The triplet "molecular processes–product–process" engineering: the future of chemical engineering?', *Chemical Engineering Science*, 57, pp. 4667–4690.

Chen, C. K.-M., Lee, G. C., Ko, T. P., Guo, R. T., Huang, L. M., Liu, H. J., Ho, Y. F., Shaw, J. F. and Wang, A. H. J. (2009) 'Structure of the Alkalohyperthermophilic *Archaeoglobus fulgidus* Lipase Contains a Unique C-Terminal Domain Essential for Long-Chain Substrate Binding', *Journal of Molecular Biology*. Elsevier B.V., 390(4), pp. 672–685.

Cheng, A. A. and Lu, T. K. (2012) 'Synthetic biology: an emerging engineering discipline.', *Annual review of biomedical engineering*, 14, pp. 155–78.

Cherubini, F. (2010) 'The biorefinery concept: Using biomass instead of oil for producing energy and chemicals', *Energy Conversion and Management*. Elsevier Ltd, 51(7), pp. 1412–1421.

Coleman, J. R., Papamichail, D., Yano, M., Del Mar García-Suárez, M. and Pirofski, L. A. (2011) 'Designed reduction of *Streptococcus pneumoniae* pathogenicity via synthetic changes in virulence factor codon-pair bias', *Journal of Infectious Diseases*, 203(9), pp. 1264–1273.

Cook, P. F. and Cleland, W. W. (2007) *Enzyme Kinetics and Mechanism*. New York: Garland Science.

Cooling, M. T., Rouilly, V., Misirli, G., Lawson, J., Yu, T., Hallinan, J. and Wipat, A. (2010) 'Standard virtual biological parts: A repository of modular modeling components for synthetic biology', *Bioinformatics*, 26(7), pp. 925–931.

Copeland, R. A. (2013) *Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists*. John Wiley & Sons.

Cowtan, K. (2010) 'Recent developments in classical density modification', *Acta Crystallographica Section D: Biological Crystallography*, 66(4), pp. 470–478.

Dalby, P. A. (2007) 'Engineering enzymes for biocatalysis', *Recent Patents on Biotechnology*, 1(1), pp. 1–9.

Damborsky, J. and Brezovsky, J. (2014) 'Computational tools for designing and engineering enzymes', *Current Opinion in Chemical Biology*. Elsevier Ltd, 19(1), pp. 8–16.

Deb, K., Agrawal, S., Pratap, A. and Meyarivan, T. (2000) 'A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II', *Parallel Problem Solving from Nature PPSN VI*, pp. 849–858.

Demirtas, M. U., Kolhatkar, A. and Kilbane, J. J. (2003) 'Effect of aeration and agitation on growth rate of *Thermus thermophilus* in batch mode.', *Journal of bioscience and bioengineering*, 95(2), pp. 113–7.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M. and Gascuel, O. (2008) 'Phylogeny.fr: robust phylogenetic analysis for the non-specialist.', *Nucleic acids research*, 36(Web Server issue), pp. 465–469.

DeSantis, G., Zhu, Z., Greenberg, W. A., Wong, K., Chaplin, J., Hanson, S. R., Farwell, B., Nicholson, L. W., Rand, C. L., Weiner, D. P., Robertson, D. E. and Burk, M. J. (2002) 'An enzyme library approach to biocatalysis: Development of nitrilases for enantioselective production of carboxylic acid derivatives', *Journal of the American Chemical Society*, 124(31), pp. 9024–9025.

Dillon, M. M., Sung, W., Lynch, M., Cooper, V. S. and Cooper, V. (2015) 'The rate and molecular spectrum of spontaneous mutations in the GC-rich multi-chromosome genome of *Burkholderia cenocepacia*', *Genetics: Early Online*, 12115(July), pp. 935–946.

Do, P. M., Angerhofer, A., Hrdy, I., Bardonova, L., Ingram, L. O. and Shanmugam, K. T. (2009) 'Engineering *Escherichia coli* for fermentative dihydrogen production: Potential role of NADH-

ferredoxin oxidoreductase from the hydrogenosome of anaerobic protozoa', *Applied Biochemistry and Biotechnology*, 153(1–3), pp. 21–33.

Domínguez de María, P. and Hollmann, F. (2015) 'On the (Un)greenness of Biocatalysis: Some Challenging Figures and Some Promising Options', *Frontiers in Microbiology*, 6(November), pp. 6–10.

Van Der Donk, W. A. and Zhao, H. (2003) 'Recent developments in pyridine nucleotide regeneration', *Current Opinion in Biotechnology*, 14(4), pp. 421–426.

Van Dover, C. L. (2000) *The Ecology of Deep-Sea Hydrothermal Vents*, *Ecology*.

Duetz, W. A., Van Beilen, J. B. and Witholt, B. (2001) 'Using proteins in their natural environment: Potential and limitations of microbial whole-cell hydroxylations in applied biocatalysis', *Current Opinion in Biotechnology*, 12(4), pp. 419–425.

Dunn, G., Montgomery, M. G., Mohammed, F., Coker, A., Cooper, J. B., Robertson, T., Garcia, J. L., Bugg, T. D. H. and Wood, S. P. (2005) 'The structure of the C-C bond hydrolase MhpC provides insights into its catalytic mechanism', *Journal of Molecular Biology*, 346(1), pp. 253–265.

Edmondson, D. E., Mattevi, A., Binda, C., Li, M., Hubálek, F. and Abraham, D. J. (2003) 'Structure and Mechanism of Monoamine Oxidase', in *Burger's Medicinal Chemistry and Drug Discovery*. John Wiley & Sons, Inc.

Ehmann, D. E., Gehring, A. M. and Walsh, C. T. (1999) 'Lysine biosynthesis in *Saccharomyces cerevisiae*: Mechanism of alpha- aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5', *Biochemistry*, 38(19), pp. 6171–6177.

Ellis, T., Adie, T. and Baldwin, G. S. (2011) 'DNA assembly for synthetic biology: from parts to pathways and beyond.', *Integrative biology : quantitative biosciences from nano to macro*, 3(2), pp. 109–118.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) 'Features and development of Coot', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 66(4), pp. 486–501.

Erickson, B., Nelson and Winters, P. (2012) 'Perspective on opportunities in industrial biotechnology in renewable chemicals.', *Biotechnology journal*, 7(2), pp. 176–85.

Espah Borujeni, A., Channarasappa, A. S. and Salis, H. M. (2014) 'Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites.', *Nucleic acids research*, 42(4), pp. 2646–59.

Estell, D. A., Graycar, T. P. and Wells, J. A. (1985) 'Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation', *Journal of Biological Chemistry*, 260(11), pp. 6518–6521.

Evans, P. R. and Murshudov, G. N. (2013) 'How good are my data and what is the resolution?', *Acta Crystallographica Section D: Biological Crystallography*, 69(7), pp. 1204–1214.

Faraldo, M. M., De Pedro, M. A., Berenguer, J., Pedro, M. A. D. E. and Berenguer, J. (1992) 'Sequence of the S-layer gene of *Thermus thermophilus* HB8 and functionality of its promoter in *Escherichia coli*', *Journal of Bacteriology*, 174(22), pp. 7458–7462.

Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M. and Salis, H. M. (2014) 'Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria.', *Molecular systems biology*, 10, p. 731.

Farkas, J., Stirrett, K., Lipscomb, G. L., Nixon, W., Scott, R. A., Adams, M. W. W. and Westpheling, J. (2012) 'Recombinogenic properties of *Pyrococcus furiosus* strain COM1 enable rapid selection of targeted mutants', *Applied and Environmental Microbiology*, 78(13), pp. 4669–4676.

Farmer, J. (1998) 'Thermophiles, early biosphere evolution, and the origin of life on Earth:

Implications for the exobiological exploration of Mars', *Journal of Geophysical Research*, 103(98), pp. 457–461.

Fernández-Herrero, L. A., Olabarria, G. and Berenguer, J. (1997) 'Surface proteins and a novel transcription factor regulate the expression of the S-layer gene in *Thermus thermophilus* HB8.', *Molecular microbiology*, 24, pp. 61–72.

Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V. and Golyshin, P. N. (2016) 'Biodiversity for biocatalysis: A review of the  $\alpha/\beta$ -hydrolase fold superfamily of esterases-lipases discovered in metagenomes', *Biocatalysis and Biotransformation*, 2422(March), pp. 1–15.

Fessner, W. D. (2015) 'Systems Biocatalysis: Development and engineering of cell-free "artificial metabolisms" for preparative multi-enzymatic synthesis', *New Biotechnology*. Elsevier B.V., 32(6), pp. 658–664.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J. and Punta, M. (2014) 'Pfam: The protein families database', *Nucleic Acids Research*, 42(D1), pp. 1–9.

Finnigan, W., Thomas, A., Cromas, H., Gough, B., Snajdrova, R., Adams, J., Littlechild, J. A. and Harmer, N. (2016) 'Characterization of carboxylic acid reductases as enzymes in the toolbox for synthetic chemistry. (Manuscript in submission)', *ChemCatChem*.

Fischer, E. (1890) 'Ueber die optischen Isomeren des Traubenzuckers, der Gluconsäure und der Zuckersäure', *Berichte der deutschen chemischen Gesellschaft*, 23(2), pp. 2611–2624.

Fleming, T. and Littlechild, J. (1997) 'Sequence and structural comparison of thermophilic phosphoglycerate kinases with a mesophilic equivalent', *Comparative Biochemistry and Physiology - A Physiology*, 118(3), pp. 439–451.

France, S. P., Hussain, S., Hill, A. M., Hepworth, L. J., Howard, R. M., Mulholland, K. R., Flitsch, S. L. and Turner, N. J. (2016) 'One Pot Cascade Synthesis of Mono- and Di-Substituted Piperidines and Pyrrolidines using Carboxylic Acid Reductase (CAR),  $\omega$ -Transaminase ( $\omega$ -TA) and Imine Reductase (IREd) Biocatalysts', *ACS Catalysis*, p. acscatal.6b00855.

French, C. E. (2009) 'Synthetic biology and biomass conversion: a match made in heaven?', *Journal of the Royal Society, Interface / the Royal Society*, 6 Suppl 4(May), pp. S547–S558.

Frock, A. D. and Kelly, R. M. (2012) 'Extreme thermophiles: Moving beyond single-enzyme biocatalysis', *Current Opinion in Chemical Engineering*. Elsevier Ltd, 1(4), pp. 363–372.

Galtier, N., Tourasse, N. and Gouy, M. (1999) 'A nonhyperthermophilic common ancestor to extant life forms.', *Science*, 283(5399), pp. 220–221.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) 'Protein Identification and Analysis Tools on the ExPASy Server', in Walker, J. M. (ed.) *The Proteomics Protocols Handbook*. 1st edn. Totowa, NJ, pp. 571–607.

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., Smith, H. O. and Venter, J. C. (2010) 'Creation of a bacterial cell controlled by a chemically synthesized genome.', *Science*, 329(5987), pp. 52–56.

Gingold, H. and Pilpel, Y. (2011) 'Determinants of translation efficiency and accuracy.', *Molecular systems biology*, 7(481), p. 481.

Girard, J. E. (1980) 'Practical organic chemistry', *Nature*, 284(5751), pp. 83–83.

Glansdorff, N. and Xu, Y. (2004) 'Phylogeny of extremophiles', in *Extremophiles*.

Goldberg, K., Schroer, K., Lütz, S. and Liese, A. (2007) 'Biocatalytic ketone reduction - A powerful

tool for the production of chiral alcohols - Part II: Whole-cell reductions', *Applied Microbiology and Biotechnology*, 76(2), pp. 249–255.

Gomes, J. and Steiner, W. (2004) 'The biocatalytic potential of extremophiles and extremozymes', *Food Technology and Biotechnology*, 42(4), pp. 223–235.

Goodrich, A. C., Harden, B. J. and Frueh, D. P. (2015) 'Solution Structure of a Nonribosomal Peptide Synthetase Carrier Protein Loaded with Its Substrate Reveals Transient, Well-Defined Contacts', *Journal of the American Chemical Society*, 137(37), pp. 12100–12109.

Gould, N., Hendy, O. and Papamichail, D. (2014) 'Computational tools and algorithms for designing customized synthetic genes.', *Frontiers in bioengineering and biotechnology*, 2, p. 41.

de Grado, M. De and Casta, P. (1999) 'A High-Transformation-Efficiency Cloning Vector for *Thermus thermophilus*', 245, pp. 241–245.

De Grado, M., Lasa, I. and Berenguer, J. (1998) 'Characterization of a plasmid replicative origin from an extreme thermophile', *FEMS Microbiology Letters*, 165(1), pp. 51–57.

Green, A. P., Turner, N. J. and O'Reilly, E. (2014) 'Chiral amine synthesis using w-transaminases: An amine donor that displaces equilibria and enables high-throughput screening', *Angewandte Chemie - International Edition*, 53(40), pp. 10714–10717.

Gross, G. G. and Zenk, M. H. (1969) 'Reduktion aromatischer Säuren zu Aldehyden und Alkoholen im zellfreien System', *European journal of biochemistry / FEBS*, 8, pp. 413–419.

Guazzaroni, M. E., Silva-Rocha, R. and Ward, R. J. (2015) 'Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening', *Microbial Biotechnology*, 8(1), pp. 52–64.

Gulick, A. M. (2009) 'Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase', *ACS Chemical Biology*, 4(10), pp. 811–827.

Guy, J. E., Isupov, M. N. and Littlechild, J. A. (2003) 'The structure of an alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix*', *Journal of Molecular Biology*, 331(5), pp. 1041–1051.

Hansch, C., Rockwell, S. D., Jow, P. Y. C., Leo, A. and Steller, E. E. (1977) 'Substituent constants for correlation analysis', *Journal of Medicinal Chemistry*, 20(2), pp. 304–306.

Hansen, E. H., Møller, B. L., Kock, G. R., Bünner, C. M., Kristensen, C., Jensen, O. R., Okkels, F. T., Olsen, C. E., Motawia, M. S. and Hansen, J. (2009) 'De novo biosynthesis of Vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*)', *Applied and Environmental Microbiology*, 75(9), pp. 2765–2774.

He, A., Li, T., Daniels, L., Fotheringham, I. and Rosazza, J. P. N. (2004) 'Nocardia sp. Carboxylic Acid Reductase: Cloning, Expression, and Characterization of a New Aldehyde Oxidoreductase Family', *Applied and Environmental Microbiology*, 70(3), pp. 1874–1881.

Heap, J. T., Pennington, O. J., Cartman, S. T. and Minton, N. P. (2009) 'A modular system for *Clostridium* shuttle plasmids', *Journal of Microbiological Methods*, 78(1), pp. 79–85.

Heider, J., Ma, K. and Adams, M. W. W. (1995) 'Purification, characterization, and metabolic function of tungsten-containing aldehyde ferredoxin oxidoreductase from the hyperthermophilic and proteolytic archaeon *Thermococcus* strain ES-1', *Journal of Bacteriology*, 177(16), pp. 4757–4764.

Heine, M. and Chandra, S. B. C. (2009) 'The linkage between reverse gyrase and hyperthermophiles: A review of their invariable association', *Journal of Microbiology*, 47(3), pp. 229–234.

Heinemann, M. and Panke, S. (2006) 'Synthetic biology - Putting engineering into biology', *Bioinformatics*, 22(22), pp. 2790–2799.

Heinonen, J. K. and Lahti, R. J. (1981) 'A new and convenient colorimetric determination of inorganic orthophosphate and its application to the assay of inorganic pyrophosphatase', *Analytical*

*Biochemistry*, 113(2), pp. 313–317.

Henne, A., Brüggemann, H., Raasch, C., Wiezer, A., Hartsch, T., Liesegang, H., Johann, A., Lienard, T., Gohl, O., Martinez-Arias, R., Jacobi, C., Starkuviene, V., Schlenczeck, S., Dencker, S., Huber, R., Klenk, H.-P., Kramer, W., Merkl, R., Gottschalk, G. and Fritz, H.-J. (2004) 'The genome sequence of the extreme thermophile *Thermus thermophilus*', *Nature biotechnology*, 22(5), pp. 547–53.

Hickey, A. M. (2008) *Unpublished work*.

Hickey, A. M., Marle, L., McCreedy, T., Watts, P., Greenway, G. M. and Littlechild, J. A. (2007) 'Immobilization of thermophilic enzymes in miniaturized flow reactors', *Biochemical Society transactions*, 35(6), pp. 1621–1623.

Hickey, A. M., Ngamsom, B., Wiles, C., Greenway, G. M., Watts, P. and Littlechild, J. A. (2009) 'A microreactor for the study of biotransformations by a cross-linked gamma-lactamase enzyme', *Biotechnology Journal*, 4(4), pp. 510–516.

Hidalgo, A., Betancor, L., Moreno, R., Zafra, O., Cava, F., Fernández-Lafuente, R., Guisán, J. M. and Berenguer, J. (2004) 'Thermus thermophilus as a cell factory for the production of a thermophilic Mn-dependent catalase which fails to be synthesized in an active form in *Escherichia coli*', *Applied and Environmental Microbiology*, 70(7), pp. 3839–3844.

Hirakawa, H., Kamiya, N. and Kwarabayashi, Y. (2004) 'Properties of an Alcohol Dehydrogenase from the Hyperthermophilic Archaeon *Aeropyrum pernix* K1', 97(3), pp. 202–206.

Hirst, C. D. (2014) *Automated BioPart characterisation for synthetic biology*. Imperial College London.

Hisanaga, Y., Ago, H., Hamada, K., Ida, K., Yamamoto, M., Hori, T., Sugahara, M., Kuramitsu, S., Yokoyama, S. and Miyano, M. (2004) 'Structural Basis of the Substrate-specific Two-step Catalysis of Long Chain Fatty Acyl-CoA Synthetase Dimer', *The Journal of biological chemistry*, 279(30), pp. 31717–26.

Hofmann, B., Tölzer, S., Pelletier, I., Altenbuchner, J., van Pée, K. H. and Hecht, H. J. (1998) 'Structural investigation of the cofactor-free chloroperoxidases.', *Journal of molecular biology*, 279(4), pp. 889–900.

Hold, C. and Panke, S. (2009) 'Towards the engineering of in vitro systems', *Journal of the Royal Society, Interface*, 6, pp. 507–521.

Holland, J. H. (1975) *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.*, U Michigan Press. Oxford. England.

Hollmann, F., Ni, Y., Hagedoorn, P. L., Arends, I. W. C. E. and Xu, J.-H. (2012) 'A biocatalytic hydrogenation of carboxylic acids', *Chemical Communications*, (207890), pp. 12056–12059.

Höllrigl, V., Hollmann, F., Kleeb, A. C., Buehler, K. and Schmid, A. (2008) 'TADH, the thermostable alcohol dehydrogenase from *Thermus* sp. ATN1: A versatile new biocatalyst for organic synthesis', *Applied Microbiology and Biotechnology*, 81(2), pp. 263–273.

Holm, L. and Rosenstrom, P. (2010) 'Dali server: Conservation mapping in 3D', *Nucleic Acids Research*, 38, pp. 545–549.

Host, G., Martensson, L.-G. and Jonsson, B. H. (2006) 'Redesign of human carbonic anhydrase II for increased esterase activity and specificity towards esters with long acyl chains', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1764(10), pp. 1601–1606.

Hsu, P. D., Lander, E. S. and Zhang, F. (2014) 'Development and applications of CRISPR-Cas9 for genome engineering', *Cell*. Elsevier, 157(6), pp. 1262–1278.

Hurst, L. D. and Merchant, A. R. (2001) 'High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes.', *Proceedings of the Royal Society B: Biological Sciences*, 268(1466), pp. 493–7.

- Van Hylckama Vlieg, J. E. T., Tang, L., Lutje Spelberg, J. H., Smilda, T., Poelarends, G. J., Bosma, T., Van Merode, A. E. J., Fraaije, M. W. and Janssen, D. B. (2001) 'Halohydrin dehalogenases are structurally and mechanistically related to short-chain dehydrogenases/reductases', *Journal of Bacteriology*, 183(17), pp. 5058–5066.
- Hyndman, D., Bauman, D. R., Heredia, V. V. and Penning, T. M. (2003) 'The aldo-keto reductase superfamily homepage', *Chemico-Biological Interactions*, 143–144(16), pp. 621–631.
- iGEM website for part BBa\_I13521 (2016). Available at: [http://parts.igem.org/Part:BBa\\_I13521:Experience](http://parts.igem.org/Part:BBa_I13521:Experience) (Accessed: 7 September 2016).
- Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matsuura, Y. and Oshima, T. (1991) 'Three-Dimensional Structure of a Highly Thermostable Enzyme, 3-Isopropylmalate Dehydrogenase of *Thermus Thermophilus* at 2.2 Å Resolution', *Journal of Molecular Biology*, 222(3), pp. 725–738.
- Isupov, M. N., Fleming, T. M., Dalby, A. R., Crowhurst, G. S., Bourne, P. C. and Littlechild, J. A. (1999) 'Crystal structure of the glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*', *Journal of Molecular Biology*, 291(3), pp. 651–660.
- Itoh, H. and Shiba, T. (2004) 'Polyphosphate synthetic activity of polyphosphate:AMP phosphotransferase in *Acinetobacter johnsonii* 210A', *Journal of Bacteriology*, 186(15), pp. 5178–5181.
- Jacob, F. and Monod, J. (1961) 'Genetic regulatory mechanisms in the synthesis of proteins', *Journal of Molecular Biology*, 3(3), pp. 318–356.
- Jeschek, M., Gerngross, D. and Panke, S. (2016) 'Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort.', *Nature communications*, 7.
- Joosten, R. P., Long, F., Murshudov, G. N. and Perrakis, A. (2014) 'The PDB\_REDO server for macromolecular structure model optimization', *IUCrJ*. International Union of Crystallography, 1, pp. 213–220.
- Kaberlin, V. R. and Bläsi, U. (2006) 'Translation initiation and the fate of bacterial mRNAs', *FEMS Microbiology Reviews*, 30(6), pp. 967–979.
- Kabsch, W. (2010) 'XDS', *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp. 125–132.
- Kallio, P., Pasztor, A., Thiel, K., Akhtar, M. K. and Jones, P. R. (2014) 'An engineered pathway for the biosynthesis of renewable propane', *Nature Communications*, 5(4731), p. 4731.
- Kaluzna, I. A., David Rozzell, J. and Kambourakis, S. (2005) 'Ketoreductases: Stereoselective catalysts for the facile synthesis of chiral alcohols', *Tetrahedron Asymmetry*, 16(22), pp. 3682–3689.
- Kameda, A., Shiba, T., Kawazoe, Y., Satoh, Y., Ihara, Y., Munekata, M., Ishige, K. and Noguchi, T. (2001) 'A novel ATP regeneration system using polyphosphate-AMP phosphotransferase and polyphosphate kinase', *Journal of Bioscience and Bioengineering*, 91(6), pp. 557–563.
- Kananavičiute, R. and Čitavičius, D. (2015) 'Genetic engineering of *Geobacillus* spp.', *Journal of Microbiological Methods*, 111, pp. 31–39.
- Kankare, J., S.Neal, G., Salminen, T., Glumhoff, T., S.Cooperman, B., Lahti, R. and Goldman, A. (1994) 'The structure of *E. coli* soluble inorganic pyrophosphatase at 2.7 Å resolution', *Protein Engineering*, 7(7), pp. 823–830.
- Kashefi, K. and Lovley, D. R. (2003) 'Extending the upper temperature limit for life.', *Science*, 301(5635), p. 934.
- Kato, N., Joung, E.-H., Yang, H.-C., Masuda, M., Shimao, M. and Yanase, H. (1991) 'Purification and characterization of aromatic acid reductase from *Nocardia asteroides* JCM 3016.', *Agricultural and*



*Biological Chemistry*, 55(3), pp. 757–762.

Kayser, K. J., Kwak, J. H., Park, H. S. and Kilbane, J. J. (2001) 'Inducible and constitutive expression using new plasmid and integrative expression vectors for *Thermus* sp.', *Letters in applied microbiology*, 32(6), pp. 412–8.

Kayser, K. and Kilbane, J. (2001) 'New host-vector system for *Thermus* spp. based on the malate dehydrogenase gene', *Journal of bacteriology*, 183(5), pp. 1–5.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012) 'Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, 28(12), pp. 1647–1649.

Keasling, J. D. (2012) 'Synthetic biology and the development of tools for metabolic engineering', *Metabolic Engineering*, 14(3), pp. 189–195.

Kennedy, M. and Krouse, D. (1999) 'Strategies for improving fermentation medium performance: a review', *Journal of Industrial Microbiology and Biotechnology*, 23(6), pp. 456–475.

Kertes, A. S., King, C. J. and Blanch, H. W. (2009) 'Extraction chemistry of fermentation product carboxylic acids', *Biotechnology and bioengineering*, 103(3), pp. 431–445.

Kim, J., Kim, S., Yoon, S., Hong, E. and Ryu, Y. (2015) 'Improved enantioselectivity of thermostable esterase from *Archaeoglobus fulgidus* toward (S)-ketoprofen ethyl ester by directed evolution and characterization of mutant esterases', *Applied Microbiology and Biotechnology*.

Kim, S., Lee, W. and Ryu, Y. (2008) 'Cloning and characterization of thermostable esterase from *Archaeoglobus fulgidus*', *Journal of Microbiology*, 46(1), pp. 100–107.

King, E. L. and Altman, C. (1956) 'A Schematic Method of Deriving the Rate Laws for Enzyme-Catalyzed Reactions', *Journal of Physical Chemistry*, 60(10), pp. 1375–1378.

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., Loftus, B., Venter, J. C. and et al. (1997) 'The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*', *Nature*, 390(6658), pp. 364–370.

Klibanov, A. M. (2001) 'Improving enzymes by using them in organic Solvents', *Nature*, 409, pp. 9–11.

Knoll, M. (2008) 'The medium-chain dehydrogenase/reductase engineering database: A systematic analysis of a diverse protein family to understand sequence–structure–function relationship', *Protein Science*, 17, pp. 1689–1697.

Kobayashi, H., Kuwae, A., Maseda, H., Nakamura, A. and Hoshino, T. (2005) 'Isolation of a low-molecular-weight, multicopy plasmid, pNHK101, from *Thermus* sp. TK10 and its use as an expression vector for *T. thermophilus* HB27', *Plasmid*, 54(1), pp. 70–79.

Koga, Y. (2012) 'Thermal adaptation of the archaeal and bacterial lipid membranes', *Archaea*, 2012.

Konak, A., Coit, D. W. and Smith, A. E. (2006) 'Multi-objective optimization using genetic algorithms: A tutorial', *Reliability Engineering and System Safety*, 91(9), pp. 992–1007.

Korkhin, Y., Kalb (Gilboa), J. A., Peretz, M., Bogin, O., Burstein, Y. and Frolov, F. (1999) 'Oligomeric integrity - The structural key to thermal stability in bacterial alcohol dehydrogenases', *Protein science*, 8, pp. 1241–1249.

Koshland, D. E. (1995) 'The Key–Lock Theory and the Induced Fit Theory', *Angewandte Chemie International Edition in English*, 33(2324), pp. 2375–2378.

Koyama, Y., Hoshino, T., Tomizuka, N. and Furukawa, K. (1986) 'Genetic transformation of the

extreme thermophile *Thermus thermophilus* and of other *Thermus* spp.', *Journal of Bacteriology*, 166(1), pp. 338–340.

Kragl, U., Kruse, W., Hummel, W. and Wandrey, C. (1996) 'Enzyme engineering aspects of biocatalysis: Cofactor regeneration as example', *Biotechnology and Bioengineering*, 52(2), pp. 309–319.

Kraut, J. (1977) 'Serine proteases: Structure and mechanism of catalysis', *Annual review of biochemistry*, 46, pp. 331–358.

Kroutil, W., Mang, H., Edegger, K. and Faber, K. (2004) 'Recent advances in the biocatalytic reduction of ketones and oxidation of sec-alcohols', *Current Opinion in Chemical Biology*, 8(2), pp. 120–126.

Kuby, S. A. (1991) *A study of enzymes*. CRC Press.

Kuhn, D., Blank, L. M., Schmid, A. and Bühler, B. (2010) 'Systems biotechnology - Rational whole-cell biocatalyst and bioprocess design', *Engineering in Life Sciences*, 10(5), pp. 384–397.

Kurz, M. (2008) 'Compatible solute influence on nucleic acids: many questions but few answers.', *Saline systems*, 4(6).

Lack, N. A., Yam, K. C., Lowe, E. E., Horsman, G. P., Owen, R. L., Sim, E. and Eltis, L. D. (2010) 'Characterization of a carbon-carbon hydrolase from *Mycobacterium tuberculosis* involved in cholesterol metabolism', *Journal of Biological Chemistry*, 285(1), pp. 434–443.

Lack, N., Lowe, E. D., Liu, J., Eltis, L. D., Noble, M. E. M., Sim, E. and Westwood, I. M. (2007) 'Structure of HsaD, a steroid-degrading hydrolase, from *Mycobacterium tuberculosis*', *Protein structure communications*. International Union of Crystallography, 64(1), pp. 2–7.

Ladkau, N., Schmid, A. and Bühler, B. (2014) 'The microbial cell-functional unit for energy dependent multistep biocatalysis.', *Current opinion in biotechnology*, 30, pp. 178–189.

Lahti, R. (1983) 'Microbial inorganic pyrophosphatases.', *Microbiological reviews*, 47(2), pp. 169–178.

Laidler, K. J. (1984) 'The development of the Arrhenius equation', *Journal of Chemical Education*, 61, pp. 494–498.

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S. and Ussery, D. W. (2015) 'Insights from 20 years of bacterial genome sequencing.', *Functional & integrative genomics*, 15(2), pp. 141–61.

Langille, M., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J., Clemente, J., Burkepile, D., Vega Thurber, R., Knight, R., Beiko, R. and Huttenhower, C. (2013) 'Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.', *Nature biotechnology*, 31(9), pp. 814–21.

Lasa, I., Castón, J. R., Fernández-Herrero, L. a, de Pedro, M. a and Berenguer, J. (1992) 'Insertional mutagenesis in the extreme thermophilic eubacteria *Thermus thermophilus* HB8.', *Molecular microbiology*, 6(11), pp. 1555–64.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) 'PROCHECK: a program to check the stereochemical quality of protein structures', *Journal of Applied Crystallography*. International Union of Crystallography, 26(November), pp. 283–291.

Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A. and Murshudov, G. N. (2012) 'JLigand: A graphical tool for the CCP4 template-restraint library', *Acta Crystallographica Section D Biological Crystallography*, 68(4), pp. 431–440.

Lee, N.-R., Lakshmanan, M., Aggarwal, S., Song, J.-W., Karimi, I. a, Lee, D.-Y. and Park, J.-B. (2014) 'Genome-scale metabolic network reconstruction and in silico flux analysis of the thermophilic bacterium *Thermus thermophilus* HB27.', *Microbial cell factories*, 13(61).

- Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P. and Chatonnet, A. (2013) 'ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins: Tools to explore diversity of functions', *Nucleic Acids Research*, 41(1), pp. 423–429.
- Leuchs, S., Na'amnieh, S. and Greiner, L. (2013) 'Enantioselective reduction of sparingly water-soluble ketones: continuous process and recycle of the aqueous buffer system', *Green Chemistry*, 15(1), pp. 167–176.
- Levisson, M., Han, G. W., Deller, M. C., Xu, Q., Biely, P., Hendriks, S., Ten Eyck, L. F., Flensburg, C., Roversi, P., Miller, M. D., McMullan, D., von Delft, F., Kreusch, A., Deacon, A. M., van der Oost, J., Lesley, S. A., Elsiger, M.-A., Kengen, S. W. M. and Wilson, I. A. (2012) 'Functional and structural characterization of a thermostable acetyl esterase from *Thermotoga maritima*', *Proteins: Structure, Function, and Bioinformatics*, 80(6), pp. 1545–1559.
- Levisson, M., van der Oost, J. and Kengen, S. W. M. (2007) 'Characterization and structural modeling of a new type of thermostable esterase from *Thermotoga maritima*.', *The FEBS journal*, 274(11), pp. 2832–2842.
- Levisson, M., van der Oost, J. and Kengen, S. W. M. (2009) 'Carboxylic ester hydrolases from hyperthermophiles', *Extremophiles*, 13(4), pp. 567–581.
- Li, G.-W., Oh, E. and Weissman, J. S. (2012) 'The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria', *Nature*. Nature Publishing Group, 484(7395), pp. 538–541.
- Li, H., Melton, E. M., Quackenbush, S., DiRusso, C. C. and Black, P. N. (2007) 'Mechanistic studies of the long chain acyl-CoA synthetase Faa1p from *Saccharomyces cerevisiae*', *Biochimica et Biophysica Acta*, 1771(9), pp. 1246–1253.
- Li, T. and Rosazza, J. P. N. (1997) 'Purification, characterization, and properties of an aryl aldehyde oxidoreductase from *Nocardia* sp. strain NRRL 5646', *Journal of Bacteriology*, 179(11), pp. 3482–3487.
- Li, W. F., Zhou, X. X. and Lu, P. (2005) 'Structural features of thermozymes', *Biotechnology Advances*, 23(4), pp. 271–281.
- Lin, J., Palomec, L. and Wheeldon, I. (2014) 'Design and Analysis of Enhanced Catalysis in Scaffolded Multienzyme Cascade Reactions', *ACS Catalysis*, 4, pp. 505–511.
- Line, K., Isupov, M. N. and Littlechild, J. A. (2004) 'The crystal structure of a (-)  $\gamma$ -lactamase from an *Aureobacterium* species reveals a tetrahedral intermediate in the active site', *Journal of Molecular Biology*, 338(3), pp. 519–532.
- Littlechild, J. A. (2011) 'Thermophilic archaeal enzymes and applications in biocatalysis.', *Biochemical Society Transactions*, 39(1), pp. 155–8.
- Littlechild, J. A. (2015) 'Enzymes from Extreme Environments and Their Industrial Applications', *Frontiers in Bioengineering and Biotechnology*, 3(October), pp. 1–9.
- Littlechild, J. a, Guy, J., Connelly, S., Mallett, L., Waddell, S., Rye, C. a, Line, K. and Isupov, M. (2007) 'Natural methods of protein stabilization: thermostable biocatalysts.', *Biochemical Society transactions*, 35(Pt 6), pp. 1558–1563.
- Littlechild, J., Novak, H., James, P. and Sayer, C. (2013) 'Mechanisms of thermal stability adopted by thermophilic proteins and their use in white biotechnology', in Satyanarayana, T., Littlechild, J., and Kawarabayasi, Y. (eds) *Thermophilic Microbes in Environmental and Industrial Biotechnology*. 2nd edn. Dordrecht: Springer Science and Business Media, pp. 481–507.
- Liu, P., Ewis, H. E., Tai, P. C., Lu, C. D. and Weber, I. T. (2007) 'Crystal Structure of the *Geobacillus stearothermophilus* Carboxylesterase Est55 and Its Activation of Prodrug CPT-11', *Journal of Molecular Biology*, 367(1), pp. 212–223.
- Liu, P., Wang, Y. F., Ewis, H. E., Abdelal, A. T., Lu, C. D., Harrison, R. W. and Weber, I. T. (2004)

- 'Covalent reaction intermediate revealed in crystal structure of the *Geobacillus stearothermophilus* carboxylesterase Est30', *Journal of Molecular Biology*, 342(2), pp. 551–561.
- Locey, K. J. and Lennon, J. T. (2016) 'Scaling laws predict global microbial diversity', *Proceedings of the National Academy of Sciences*, Early Edit, pp. 1–6.
- Lodders, N. and Kämpfer, P. (2007) 'Streptomycetaceae: Phylogeny, Ecology and Pathogenicity', *eLS*.
- Lomolino, G., Rizzi, C., Spettoli, P., Curioni, A. and Lante, A. (2003) 'Cell vitality and esterase activity of *Saccharomyces cerevisiae* is affected by increasing calcium concentration', *Agro Food Industry Hi Tech*. Teknosienze SRL, 14(6), pp. 32–36.
- Madigan, M. T., Martinko, J. M., Stahl, D. and Clark, D. P. (2012) *Brock Biology of Microorganisms*. 13th edn, *International Microbiology*. 13th edn. San Francisco: Benjamin Cummings.
- Mali, P., Esvelt, K. M. and Church, G. M. (2013) 'Cas9 as a versatile tool for engineering biology', *Nature Methods*, 10(10), pp. 957–963.
- Manco, G., Giosuè, E., D'auria, S., Herman, P., Carrea, G. and Rossi, M. (2000) 'Cloning, Overexpression, and properties of a new thermophilic and thermostable esterase with sequence similarity to hormone-sensitive lipase subfamily from the archaeon *Archaeoglobus fulgidus*', *Archives Biochemistry Biophysics*, 373(1), pp. 182–192.
- Mandrich, L., Menchise, V., Alterio, V., De Simone, G., Pedone, C., Rossi, M. and Manco, G. (2008) 'Functional and structural features of the oxyanion hole in a thermophilic esterase from *Alicyclobacillus acidocaldarius*', *Proteins: Structure, Function, and Bioinformatics*, 71(4), pp. 1721–1731.
- Marahiel, M. A., Stachelhaus, T. and Mootz, H. D. (1997) 'Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis.', *Chemical Reviews*, 97(7), pp. 2651–2674.
- Marashi, S. A. and Ghalanbor, Z. (2004) 'Correlations between genomic GC levels and optimal growth temperatures are not "robust"', *Biochemical and Biophysical Research Communications*, 325(2), pp. 381–383.
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C. and Bryant, S. H. (2015) 'CDD: NCBI's conserved domain database.', *Nucleic acids research*. England, 43(Database issue), pp. D222–6.
- Marguet, E. and Forterre, P. (1998) 'Protection of DNA by salts against thermodegradation at temperatures typical for hyperthermophiles', *Extremophiles*, 2(2), pp. 115–122.
- Matsumura, M. and Aiba, S. (1985) 'Screening for thermostable mutant of kanamycin nucleotidyltransferase by the use of a transformation system for a thermophile, *Bacillus stearothermophilus*', *Journal of Biological Chemistry*, 260(28), pp. 15298–15303.
- McLachlan, M. J., Johannes, T. W. and Zhao, H. (2007) 'Further Improvement of Phosphite Dehydrogenase Thermostability by Saturation Mutagenesis', *Biotechnology and Bioengineering*, 99(2), pp. 268–274.
- McNicholas, S., Potterton, E., Wilson, K. S. and Noble, M. E. M. (2011) 'Presenting your structures: The CCP4mg molecular-graphics software', *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp. 386–394.
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P. and Lopez, R. (2013) 'Analysis Tool Web Services from the EMBL-EBI.', *Nucleic acids research*, 41, pp. 597–600.
- Medema, M. H., van Raaphorst, R., Takano, E. and Breitling, R. (2012) 'Computational tools for the synthetic design of biochemical pathways.', *Nature reviews. Microbiology*, 10(3), pp. 191–202.

- Mega, R., Manzoku, M., Shinkai, A., Nakagawa, N., Kuramitsu, S. and Masui, R. (2010) 'Very rapid induction of a cold shock protein by temperature downshift in *Thermus thermophilus*', *Biochemical and Biophysical Research Communications*, 399(3), pp. 336–340.
- Mehta, D. and Satyanarayana, T. (2013) 'Diversity of hot environments and thermophilic microbes', in Satyanarayana, T., Littlechild, J., and Kawarabayasi, Y. (eds) *Thermophilic Microbes in Environmental and Industrial Biotechnology SE - 19*. 2nd edn. Dordrecht: Springer Science and Business Media, pp. 3–60.
- Mei, Y., Peng, N., Zhao, S., Hu, Y., Wang, H., Liang, Y. and She, Q. (2012) 'Exceptional thermal stability and organic solvent tolerance of an esterase expressed from a thermophilic host.', *Applied microbiology and biotechnology*, 93(5), pp. 1965–74.
- Meyer, H.-P., Eichhorn, E., Hanlon, S., Lütz, S., Schürmann, M., Wohlgemuth, R. and Coppolecchia, R. (2013) 'The use of enzymes in organic synthesis and the life sciences: perspectives from the Swiss Industrial Biocatalysis Consortium (SIBC).', *Catalysis Science & Technology*, 3, pp. 29–40.
- Meyer, H.-P., Ghisalba, O. and Leresche, J. E. (2014) 'Biotransformations and the Pharma industry', in Anastas, P. T. and Crabtree, R. H. (eds) *Handbook of Green Chemistry. Volume 3*. Wiley, pp. 171–207.
- Mitchell, C. A., Shi, C., Aldrich, C. C. and Gulick, A. M. (2012) 'Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains', *Biochemistry*, 51(15), pp. 3252–3263.
- Molbiol-Tools Web Portal* (2016). Available at: <http://molbiol-tools.ca/Promoters.htm> (Accessed: 7 September 2016).
- Moreno, R., Haro, A., Castellanos, A. and Berenguer, J. (2005) 'High-level overproduction of his-tagged Tth DNA polymerase in *Thermus thermophilus*', *Applied and Environmental Microbiology*, 71(1), pp. 591–593.
- Moreno, R., Zafra, O., Cava, F. and Berenguer, J. (2003) 'Development of a gene expression vector for *Thermus thermophilus* based on the promoter of the respiratory nitrate reductase', *Plasmid*, 49(1), pp. 2–8.
- Moura, M., Pertusi, D., Lenzi, S., Bhan, N., Broadbelt, L. J. and Tyo, K. E. J. (2015) 'Characterizing and predicting carboxylic acid reductase activity for diversifying bioaldehyde production', *Biotechnology and Bioengineering*, 113(5), pp. 944–952.
- Mu, H., Zhou, S. M., Xia, Y., Zou, H., Meng, F. and Yan, Y. Bin (2009) 'Inactivation and unfolding of the hyperthermophilic inorganic pyrophosphatase from *Thermus thermophilus* by sodium dodecyl sulfate', *International Journal of Molecular Sciences*, 10(6), pp. 2849–2859.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A. A. (2011) 'REFMAC5 for the refinement of macromolecular crystal structures', *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp. 355–367.
- Murzina, N. V., Vorozheykina, D. P. and Matvienko, N. I. (1988) 'Nucleotide sequence of *Thermus thermophilus* HB8 gene coding 16S rRNA', *Nucleic Acids Research*, 16(16), p. 8172.
- Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P. and Endy, D. (2013) 'Precise and reliable gene expression via standard transcription and translation initiation elements.', *Nature methods*, 10(4), pp. 354–60.
- Nakamura, A., Takakura, Y., Kobayashi, H. and Hoshino, T. (2005) 'In vivo directed evolution for thermostabilization of *Escherichia coli* hygromycin B phosphotransferase and the use of the gene as a selection marker in the host-vector system of *Thermus thermophilus*.', *Journal of bioscience and bioengineering*, 100(2), pp. 158–63.
- Nakamura, Y., Gojobori, T. and Ikemura, T. (1999) 'Codon usage tabulated from the international DNA sequence data bank; its status.', *Nucleic Acids Res.*, 27(1), p. 292.

- Nam, K. Y., Struck, D. K. and Holtzapple, M. T. (1996) 'ATP Regeneration by Thermostable ATP Synthase'.
- Napora-Wijata, K., Robins, K., Osorio-Lozada, A. and Winkler, M. (2014) 'Whole-Cell Carboxylate Reduction for the Synthesis of 3-Hydroxytyrosol', *ChemCatChem*, 6, pp. 1089–1085.
- Napora-Wijata, K., Strohmeier, G. A. and Winkler, M. (2014) 'Biocatalytic reduction of carboxylic acids', *Biotechnology Journal*, 9(6), pp. 822–843.
- Ni, Y. and Chen, R. R. (2004) 'Accelerating whole-cell biocatalysis by reducing outer membrane permeability barrier', *Biotechnology and Bioengineering*, 87(6), pp. 804–811.
- Ni, Y., Hagedoorn, P. L., Xu, J. H., Arends, I. W. C. E. and Hollmann, F. (2013) 'Pyrococcus furiosus-mediated reduction of conjugated carboxylic acids: Towards using syngas as reductant', *Journal of Molecular Catalysis B: Enzymatic*, 103, pp. 52–55.
- Ninh, P. H., Honda, K., Sakai, T., Okano, K. and Ohtake, H. (2015) 'Assembly and multiple gene expression of thermophilic enzymes in Escherichia coli for in vitro metabolic engineering', *Biotechnology and Bioengineering*, 112(1), pp. 189–196.
- Nossent, J., Elsen, P. and Bauwens, W. (2011) 'Sobol' sensitivity analysis of a complex environmental model', *Environmental Modelling & Software*, 26(12), pp. 1515–1525.
- Nystrom, R. F. and Brown, W. G. (1947) 'Reduction of Organic Compounds by Lithium Aluminium Hydride. II Carboxylic Acids', *Journal of the American Chemical Society*, pp. 2548–2549.
- O'Reilly, E., Köhler, V., Flitsch, S. L. and Turner, N. J. (2011) 'Cytochromes P450 as useful biocatalysts: addressing the limitations.', *Chemical communications*, 47(9), pp. 2490–501.
- Ohtani, N., Tomita, M. and Itaya, M. (2013) 'Identification of a replication initiation protein of the pVV8 plasmid from Thermus thermophilus HB8', *Extremophiles*, 17(1), pp. 15–28.
- Ohtani, N., Tomita, M. and Itaya, M. (2016) 'Curing the megaplasmid pTT27 from Thermus thermophilus HB27, and maintaining exogenous plasmids in the plasmid-free strain.', *Applied and environmental microbiology*, 82(5), pp. 1537–1548.
- Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I. and Schrag, J. (1992) 'The alpha/beta hydrolase fold.', *Protein engineering*, 5(3), pp. 197–211.
- Olson, D. G., McBride, J. E., Joe Shaw, A. and Lynd, L. R. (2012) 'Recent progress in consolidated bioprocessing', *Current Opinion in Biotechnology*, 23(3), pp. 396–405.
- Omelchenko, M. V, Wolf, Y. I., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Daly, M. J., Koonin, E. V and Makarova, K. S. (2005) 'Comparative genomics of Thermus thermophilus and Deinococcus radiodurans: divergent routes of adaptation to thermophily and radiation resistance', *BMC Evolutionary biology*, 5(57).
- Oroz-Guinea, I. and García-Junceda, E. (2013) 'Enzyme catalysed tandem reactions', *Current Opinion in Chemical Biology*, 17(2), pp. 236–249.
- Pace, C. N. (1992) 'Contribution of the hydrophobic effect to globular protein stability', *Journal of Molecular Biology*, 226(1), pp. 29–35.
- Panda, T. and Gowrishankar, B. S. (2005) 'Production and applications of esterases', *Applied Microbiology and Biotechnology*, 67(2), pp. 160–169.
- Park, G.-J. (2007) 'Design of experiments', in *Analytic Methods for Design Practice*. 1st edn. London: Springer-Verlag, pp. 309–391.
- Park, H.-S. and Kilbane, J. J. (2004) 'Gene expression studies of Thermus thermophilus promoters PdnaK, Parg and Pscs-mdh.', *Letters in applied microbiology*, 38(5), pp. 415–22.

- Peng, N., Deng, L., Mei, Y., Jiang, D., Hu, Y., Awayez, M., Liang, Y. and She, Q. (2012) 'A synthetic arabinose-inducible promoter confers high levels of recombinant protein expression in hyperthermophilic archaeon *Sulfolobus islandicus*', *Applied and Environmental Microbiology*, 78(16), pp. 5630–5637.
- Peri, S., Karra, S., Lee, Y. Y. and Karim, M. N. (2007) 'Modeling intrinsic kinetics of enzymatic cellulose hydrolysis', *Biotechnology Progress*, 23(3), pp. 626–637.
- Perrier, V., Burlacu-Miron, S., Bourgeois, S., Surewicz, W. K. and Gilles, a M. (1998) 'Genetically engineered zinc-chelating adenylate kinase from *Escherichia coli* with enhanced thermal stability.', *The Journal of biological chemistry*, 273(30), pp. 19097–19101.
- Polizzi, K. M., Bommarius, A. S., Broering, J. M. and Chaparro-Riggers, J. F. (2007) 'Stability of biocatalysts', *Current Opinion in Chemical Biology*, 11(2), pp. 220–225.
- Pollard, D. J. and Woodley, J. M. (2007) 'Biocatalysis for pharmaceutical intermediates: the future is now', *Trends in Biotechnology*, 25(2), pp. 66–73.
- du Preez, R., Clarke, K. G., Callanan, L. H. and Burton, S. G. (2015) 'Modelling of immobilised enzyme biocatalytic membrane reactor performance', *Journal of Molecular Catalysis B: Enzymatic*, 119, pp. 48–53.
- Price, J. A., Mathias, N., Woodley, J. and Kjobsted, J. (2013) 'Application of Uncertainty and Sensitivity Analysis to a Kinetic Model for Enzymatic Biodiesel Production', *Proceedings of 12th IFAC Symposium on Computer Applications in Biotechnology*, 12(1), pp. 161–168.
- Purcell, O., Jain, B., Karr, J. R., Covert, M. W. and Lu, T. K. (2013) 'Towards a whole-cell modeling approach for synthetic biology', *Chaos*, 23(2).
- Quax, W. J. and Broekhuizen, C. P. (1994) 'Development of a new *Bacillus carboxyl* esterase for use in the resolution of chiral drugs', *Applied Microbiology and Biotechnology*, 41(4), pp. 425–431.
- Raj, S. B., Ramaswamy, S. and Plapp, B. V (2014) 'Yeast Alcohol Dehydrogenase Structure and Catalysis Yeast Alcohol Dehydrogenase Structure and Catalysis', *Biochemistry*, 53, pp. 5791–5803.
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. and Barabasi, A. L. (2002) 'Hierarchical Organization of Modularity in Metabolic Networks', *Science*, 297, pp. 1551–1555.
- Reid, M. F. and Fewson, C. A. (1994) 'Molecular characterization of microbial alcohol dehydrogenases.', *Critical reviews in microbiology*, 20(1), pp. 13–56.
- Resnick, S. M. and Zehnder, A. J. B. (2000) 'In vitro ATP regeneration from polyphosphate and AMP by polyphosphate:AMP phosphotransferase and adenylate kinase from *Acinetobacter johnsonii* 210A', *Applied and Environmental Microbiology*, 66(5), pp. 2045–2051.
- Restiawaty, E., Iwasa, Y., Maya, S., Honda, K., Omasa, T., Hirota, R., Kuroda, A. and Ohtake, H. (2011) 'Feasibility of thermophilic adenosine triphosphate-regeneration system using *Thermus thermophilus* polyphosphate kinase', *Process Biochemistry*, 46(9), pp. 1747–1752.
- Ricca, E., Brucher, B. and Schrittwieser, J. H. (2011) 'Multi-enzymatic cascade reactions: Overview and perspectives', *Advanced Synthesis and Catalysis*, 353(13), pp. 2239–2262.
- Riddles, P. W., Blakeley, R. L. and Zerner, B. (1983) 'Reassessment of Ellman's reagent.', *Methods in enzymology*, 91, pp. 49–60.
- Ringborg, R. H. and Woodley, J. M. (2016) 'The application of reaction engineering to biocatalysis', *Reaction Chemistry and Engineering*. Royal Society of Chemistry, 1, pp. 10–22.
- Rios-Solis, L., Morris, P., Grant, C., Odeleye, A. O. O., Hailes, H. C., Ward, J. M., Dalby, P. A., Baganz, F. and Lye, G. J. (2015) 'Modelling and optimisation of the one-pot, multi-enzymatic synthesis of chiral amino-alcohols based on microscale kinetic parameter determination', *Chemical Engineering Science*, 122, pp. 360–372.

- Robins, K. and Osorio-Lozada, A. (2012) 'Exploiting duality in nature: industrial examples of enzymatic oxidation and reduction reactions', *Catalysis Science & Technology*, 2(8), pp. 1524–1530.
- Robinson, H., Gao, Y. G., McCrary, B. S., Edmondson, S. P., Shriver, J. W. and Wang, A. H. (1998) 'The hyperthermophile chromosomal protein Sac7d sharply kinks DNA', *Nature*, 392(6672), pp. 202–5.
- Rogers, T. A. and Bommarius, A. S. (2011) 'Utilizing simple biochemical measurements to predict lifetime output of biocatalysts in continuous isothermal processes', *Chemical Engineering Science*, 65(6), pp. 2118–2124.
- Rollié, S., Mangold, M. and Sundmacher, K. (2012) 'Designing biological systems: Systems Engineering meets Synthetic Biology', *Chemical Engineering Science*, 69(1), pp. 1–29.
- Romano, D., Bonomi, F., Carlos De Mattos, M., De Sousa Fonseca, T., Da, M., Ferreira De Oliveira, C. and Molinari, F. (2015) 'Esterases as stereoselective biocatalysts', *Biotechnology Advances*, 33(5), pp. 547–565.
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. and Huelsenbeck, J. P. (2012) 'MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space', *Systematic Biology*, 61(3), pp. 539–542.
- Roy, J. J. and Abraham, T. E. (2004) 'Strategies in making cross-linked enzyme crystals', *Chemical Reviews*, 104(9), pp. 3705–3721.
- Rozanov, A. S., Logacheva, M. D. and Peltek, S. E. (2014) 'Draft Genome Sequences of *Geobacillus stearothermophilus* Strains 22 and 53, Isolated from the Garga Hot Spring in the Barguzin River Valley of the Russian Federation.', *Genome announcements*, 2(6), pp. e01205-14.
- Sainz de Murieta, I., Bultelle, M. and Kitney, R. I. (2016) 'Towards the first data acquisition standard in Synthetic Biology', *ACS Synthetic Biology*, 5, pp. 817–826.
- Sakaki, Y. and Oshima, T. (1975) 'Isolation and characterization of a bacteriophage infectious to an extreme thermophile, *Thermus thermophilus* HB8.', *Journal of virology*, 15(6), pp. 1449–1453.
- Salis, H. M., Mirsky, E. A. and Voigt, C. A. (2009) 'Automated design of synthetic ribosome binding sites to control protein expression.', *Nature biotechnology*, 27(10), pp. 946–950.
- Sambrook, J. and Russell, D. W. (2012) *Molecular cloning: a laboratory manual*. 4th edn. New York: Cold Spring Harbour.
- Sandman, K., Pereira, S. L. and Reeve, J. N. (1998) 'Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome', *Cellular and Molecular Life Sciences*, 54(12), pp. 1350–1364.
- Sanishvili, R., Yakunin, A. F., Laskowski, R. A., Skarina, T., Evdokimova, E., Doherty-Kirby, A., Lajoie, G. A., Thornton, J. M., Arrowsmith, C. H., Savchenko, A., Joachimiak, A. and Edwards, A. M. (2003) 'Integrating structure, bioinformatics, and enzymology to discover function: BioH, a new carboxylesterase from *Escherichia coli*', *Journal of Biological Chemistry*, 278(28), pp. 26039–26045.
- Santacoloma, P. A., Sin, G., Gernaey, K. V. and Woodley, J. M. (2011) 'Multienzyme-Catalyzed Processes : Next-Generation Biocatalysis', *Organic Process Research and Development*, (5), pp. 203–212.
- Sato, M., Masuda, Y., Kirimura, K. and Kino, K. (2007) 'Thermostable ATP regeneration system using polyphosphate kinase from *Thermosynechococcus elongatus* BP-1 for D-amino acid dipeptide synthesis.', *Journal of bioscience and bioengineering*, 103(2), pp. 179–84.
- Sayer, C., Finnigan, W., Isupov, M. N., Levisson, M., Kengen, S. W. M., van der Oost, J., Harmer, N. J. and Littlechild, J. A. (2016) 'Structural and biochemical characterisation of *Archaeoglobus fulgidus* esterase reveals a bound CoA molecule in the vicinity of the active site', *Scientific Reports*, 6, pp. 25542–25555.



Sayer, C., Isupov, M. N., Bonch-Osmolovskaya, E. and Littlechild, J. A. (2015) 'Structural studies of a thermophilic esterase from a new Planctomycetes species, *Thermogutta terrifontis*', *FEBS Journal*, 282(15), pp. 2846–2857.

Schomburg, I., Chang, A., Placzek, S., Sohngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M. and Schomburg, D. (2013) 'BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA.', *Nucleic acids research*, 41, pp. 764–772.

Schrag, J. D. and Cygler, M. (1997) 'Lipases and alpha/beta hydrolase fold.', *Methods in enzymology*, 284, pp. 85–107.

Schrewe, M., Julsing, M. K., Bühler, B. and Schmid, A. (2013) 'Whole-cell biocatalysis for selective and productive C-O functional group introduction and modification.', *Chemical society reviews*, 42(15), pp. 6346–77.

Schrittwieser, J. H., Sattler, J., Resch, V., Mutti, F. G. and Kroutil, W. (2011) 'Recent biocatalytic oxidation-reduction cascades', *Current Opinion in Chemical Biology*, 15(2), pp. 249–256.

Sheldon, R. A. (2011) 'Cross-linked enzyme aggregates as stable biocatalysts', *Organic Process Research and Development*, 15(1), pp. 2694–2694.

Sheldon, R. A. and Van Pelt, S. (2013) 'Enzyme immobilisation in biocatalysis: why, what and how', *Chemical Society Reviews*, 42, pp. 6223–6235.

Sheng, X. R., Li, X. and Pan, X. M. (1999) 'An iso-random Bi Bi mechanism for adenylate kinase', *Journal of Biological Chemistry*, 274(32), pp. 22238–22242.

Shiba, T., Itoh, H., Kameda, A., Kobayashi, K., Kawazoe, Y. and Noguchi, T. (2005) 'Polyphosphate:AMP phosphotransferase as a polyphosphate-dependent nucleoside monophosphate kinase in *Acinetobacter johnsonii* 210A', *Journal of Bacteriology*, 187(5), pp. 1859–1865.

De Simone, G., Menchise, V., Manco, G., Mandrich, L., Sorrentino, N., Lang, D., Rossi, M. and Pedone, C. (2001) 'The crystal structure of a hyper-thermophilic carboxylesterase from the archaeon *Archaeoglobus fulgidus*.'', *Journal of molecular biology*, 314(3), pp. 507–18.

Sin, G., Gernaey, K. V and Lantz, A. E. (2009) 'Good modelling practice (GMoP) for PAT applications: Propagation of input uncertainty and sensitivity analysis', *Biotechnology Progress*, 25, pp. 1043–1053.

Sin, G., Woodley, J. M. and Gernaey, K. V. (2009) 'Application of modeling and simulation tools for the evaluation of biocatalytic processes: A future perspective', *Biotechnology Progress*, 25(6), pp. 1529–1538.

Singleton, M. R., Isupov, M. N. and Littlechild, J. A. (1999) 'X-ray structure of pyrrolidone carboxyl peptidase from the hyperthermophilic', *Structure*, 7(3), p. 237.

Solomatine, D. P. and Ostfeld, A. (2008) 'Data-driven modelling: some past experiences and new approaches', *Journal of Hydroinformatics*, 10(1), p. 3.

Somers, N. A. and Kazlauskas, R. J. (2004) 'Mapping the substrate selectivity and enantioselectivity of esterases from thermophiles', *Tetrahedron Asymmetry*, 15(18), pp. 2991–3004.

Stepankova, V., Damborsky, J. and Chaloupkova, R. (2013) 'Organic co-solvents affect activity, stability and enantioselectivity of haloalkane dehalogenases', *Biotechnology Journal*, 8(6), pp. 719–729.

Stetter, K. O. (1996) 'Hyperthermophiles in the history of life', *Philosophical Transactions of The Royal Society*, (361), pp. 1837–1843.

Strauss, U. P. and Siegel, A. (1963) 'The binding of magnesium ion by polyphosphates in aqueous electrolyte solutions', *Counterion Binding By Polyelectrolytes*, 67, pp. 2683–2687.

Stryer, L. (1995) *Biochemistry, Biochemistry*. New York: W.H.Freeman & Co Ltd.

- Stuermer, R., Hauer, B., Hall, M. and Faber, K. (2007) 'Asymmetric bioreduction of activated C=C bonds using enoate reductases from the old yellow enzyme family', *Current Opinion in Chemical Biology*, 11(2), pp. 203–213.
- Sundlov, J. A., Fontaine, D. M., Southworth, T. L., Branchini, B. R. and Gulick, A. M. (2012) 'Crystal structure of firefly luciferase in a second catalytic conformation supports a domain alternation mechanism', *Biochemistry*, 51(33), pp. 6493–6495.
- Tabor, S. (2001) 'Expression using the T7 RNA polymerase/promoter system.', in *Current protocols in molecular biology*. John Wiley & Sons, Inc., p. 16.2.1–16.2.11.
- Takami, H., Takaki, Y., Chee, G. J., Nishi, S., Shimamura, S., Suzuki, H., Matsui, S. and Uchiyama, I. (2004) 'Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*', *Nucleic Acids Research*, 32(21), pp. 6292–6303.
- Takayama, G., Kosuge, T., Maseda, H., Nakamura, A. and Hoshino, T. (2004) 'Nucleotide sequence of the cryptic plasmid pTT8 from *Thermus thermophilus* HB8 and isolation and characterization of its high-copy-number mutant', *Plasmid*, 51(3), pp. 227–237.
- Tamakoshi, M., Uchida, M., Tanabe, K., Fukuyama, S., Yamagishi, A. and Oshima, T. (1997) 'A new *Thermus-Escherichia coli* shuttle integration vector system', *Journal of bacteriology*, 179(15), pp. 4811–4814.
- Tamakoshi, M., Yamagishi, A. and Oshima, T. (1995) 'Screening of stable proteins in an extreme thermophile, *Thermus thermophilus*', *Molecular Microbiology*, 16(5), pp. 1031–1036.
- Tamakoshi, M., Yaoi, T., Oshima, T. and Yamagishi, a (1999) 'An efficient gene replacement and deletion system for an extreme thermophile, *Thermus thermophilus*.', *FEMS microbiology letters*, 173(2), pp. 431–437.
- Taylor, M. P., van Zyl, L., Tuffin, I. M., Leak, D. J. and Cowan, D. A. (2011) 'Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts.', *Microbial biotechnology*, 4(4), pp. 438–48.
- Tepljakov, A., Obmolova, G., Wilson, K. S., Ishii, K., Kaji, H., Samejima, T. and Kuranova, I. (1994) 'Crystal structure of inorganic pyrophosphatase from *Thermus thermophilus*.', *Protein science*, 3(7), pp. 1098–1107.
- Tishkov, V. I., Galkin, A. G., Fedorchuk, V. V., Savitsky, P. A., Rojkova, A. M., Gieren, H. and Kula, M. (1999) 'Pilot Scale Production and Isolation of Recombinant NAD<sup>+</sup> - and NADP<sup>+</sup> -Specific Formate Dehydrogenases', *Biotechnology and Bioengineering*, 64(2), pp. 187–193.
- Tomlinson, G., Mutus, B. and McLennan, I. (1981) 'Activation and inactivation of acetylcholinesterase by metal ions', *Canadian Journal of Biochemistry*, 59, pp. 728–735.
- Toogood, H. S., Brown, R. C., Line, K., Keene, P. A., Taylor, S. J. C., McCague, R. and Littlechild, J. A. (2004) 'The use of a thermostable signature amidase in the resolution of the bicyclic synthon (rac)-gamma-lactam', *Tetrahedron*, 60(3), pp. 711–716.
- Torres Pazmino, D. E., Dudek, H. M. and Fraaije, M. W. (2010) 'Baeyer-Villiger monooxygenases: recent advances and future challenges', *Current Opinion in Chemical Biology*, 14(2), pp. 138–144.
- Tufvesson, P., Lima-Ramos, J., Haque, N. Al, Gernaey, K. V. and Woodley, J. M. (2013) 'Advances in the process development of biocatalytic processes', *Organic Process Research and Development*, 17(10), pp. 1233–1238.
- Tufvesson, P., Lima-Ramos, J., Nordblad, M. and Woodley, J. M. (2011) 'Guidelines and cost analysis for catalyst production in biocatalytic processes', *Organic Process Research and Development*, 15(1), pp. 266–274.
- Tuller, T., Waldman, Y. Y., Kupiec, M. and Ruppin, E. (2010) 'Translation efficiency is determined by both codon bias and folding energy.', *Proceedings of the National Academy of Sciences of the United*

*States of America*, 107(8), pp. 3645–50.

Tullman-Ercek, D. (2015) 'Metabolism: "Channeling" Hans Krebs', *Nature Chemical Biology*, 11(3), pp. 180–181.

Turner, N. J. and O'Reilly, E. (2013) 'Biocatalytic retrosynthesis', *Nature Chemical Biology*, 9(5), pp. 285–288.

Turner, N. J. and Truppo, M. D. (2013) 'Biocatalysis enters a new era', *Current Opinion in Chemical Biology*, 17(2), pp. 212–214.

Tyagi, S. and Pleiss, J. (2006) 'Biochemical profiling in silico-Predicting substrate specificities of large enzyme families', *Journal of Biotechnology*, 124(1), pp. 108–116.

Upton, C. and Buckley, J. T. (1995) 'A new family of lipolytic enzymes?', *Trends in Biochemical Sciences*, 20(5), pp. 178–179.

Urlacher, V. B. and Girhard, M. (2012) 'Cytochrome P450 monooxygenases: An update on perspectives for synthetic application', *Trends in Biotechnology*, 30(1), pp. 26–36.

Vaguine, A. A., Richelle, J. and Wodak, S. J. (1999) 'SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model.', *Acta crystallographica. Section D, Biological crystallography*, 55, pp. 191–205.

Vasić-Rački, D., Findrik, Z. and Vrsalović Presečki, A. (2011) 'Modelling as a tool of enzyme reaction engineering for enzyme reactor development', *Applied Microbiology and Biotechnology*, 91(4), pp. 845–856.

Venkatasubramanian, P., Daniels, L. and Rosazza, J. P. N. (2007) 'Reduction of carboxylic acids by *Nocardia* aldehyde oxidoreductase requires a phosphopantetheinylated enzyme', *Journal of Biological Chemistry*, 282(1), pp. 478–485.

Vieille, C., Krishnamurthy, H., Hyun, H.-H., Savchenko, A., Yan, H. and Zeikus, J. G. (2003) 'Thermotoga neapolitana adenylate kinase is highly active at 30 degrees C.', *The Biochemical journal*, 372(Pt 2), pp. 577–85.

Vieille, C., Zeikus, G. J. and Vieille, C. (2001) 'Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability', *Microbiology and Molecular Biology reviews*, 65(1), pp. 1–43.

Vivoli, M., Novak, H. R., Littlechild, J. A. and Harmer, N. J. (2014) 'Determination of protein-ligand interactions using differential scanning fluorimetry.', *Journal of visualized experiments*, (91), p. 51809.

Voet, D. and Voet, J. G. (2004) *Biochemistry*. John Wiley & Sons.

Wachtmeister, J., Jakoblinnert, A., Kulig, J., Offermann, H. and Rother, D. (2014) 'Whole-Cell Teabag Catalysis for the Modularisation of Synthetic Enzyme Cascades in Micro-Aqueous Systems', *ChemCatChem*, 6(4), pp. 1051–1058.

Wagner, M., van Wolferen, M., Wagner, A., Lassak, K., Meyer, B. H., Reimann, J. and Albers, S. V. (2012) 'Versatile genetic tool box for the crenarchaeote *Sulfolobus acidocaldarius*', *Frontiers in Microbiology*, 3, pp. 1–12.

Wagner, U. G., Petersen, E. I., Schwab, H. and Kratky, C. (2002) 'EstB from *Burkholderia gladioli*: a novel esterase with a beta-lactamase fold reveals steric factors to discriminate between esterolytic and beta-lactam cleaving activity.', *Protein science*, 11(3), pp. 467–78.

Wei, L. L. and Goux, W. J. (1992) 'ATP cofactor regeneration via the glycolytic pathway', *Bioorganic Chemistry*, 20(1), pp. 62–66.

Wei, R., Oeser, T. and Zimmermann, W. (2014) 'Synthetic polyester-hydrolyzing enzymes from thermophilic actinomycetes', *Advances in Applied Microbiology*, 89, pp. 267–305.

- Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J. and Gustafsson, C. (2009) 'Design parameters to control synthetic gene expression in *Escherichia coli*.' *PloS one*, 4(9).
- Wells, A. and Meyer, H. P. (2014) 'Biocatalysis as a strategic green technology for the chemical industry', *ChemCatChem*, 6(4), pp. 918–920.
- Wenda, S., Illner, S., Mell, A. and Kragl, U. (2011) 'Industrial biotechnology—the future of green chemistry?', *Green Chemistry*, 13(11), p. 3007.
- Wheeldon, I., Minter, S. D., Banta, S., Barton, S. C., Atanassov, P. and Sigman, M. (2016) 'Substrate channelling as an approach to cascade reactions', *Nature Chemistry*, 8(4), pp. 299–309.
- Windle, C. L., Muller, M., Nelson, A. and Berry, A. (2014) 'Engineering aldolases as biocatalysts', *Current Opinion in Chemical Biology*, 19(1), pp. 25–33.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. and Wilson, K. S. (2011) 'Overview of the CCP4 suite and current developments', *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp. 235–242.
- Winter, G., Lobley, C. M. C. and Prince, S. M. (2013) 'Decision making in xia2', *Acta Crystallographica Section D: Biological Crystallography*, 69(7), pp. 1260–1273.
- Wohlgemuth, R. (2010) 'Biocatalysis-key to sustainable industrial chemistry', *Current Opinion in Biotechnology*, 21(6), pp. 713–724.
- Wong, C. and Whitesides, G. M. (1980) 'Enzyme-Catalyzed Organic Synthesis: NAD(P)H Cofactor Regeneration by Using Glucose 6-Phosphate and the Glucose-6-phosphate Dehydrogenase from *Leuconostoc mesenteroides*', *Journal of the American Chemical Society*, 103(8), p. 10.
- Woodley, J. and Lilly, M. D. (1990) 'Extractive biocatalysis: the use of two-liquid phase biocatalytic reactors to assist product recovery', *Chemical Engineering Science*, 45(8), pp. 2391–2396.
- Woodley, J. M. (2008) 'New opportunities for biocatalysis: making pharmaceutical processes greener', *Trends in Biotechnology*, 26(6), pp. 321–327.
- Woodley, J. M. (2013) 'Protein engineering of enzymes for process applications', *Current Opinion in Chemical Biology*. Elsevier Ltd, 17(2), pp. 310–316.
- Woodyer, R. D., Johannes, T. W. and Zhao, H. (2003) 'Regeneration of cofactors for use in biocatalysis.', in *Enzyme Technology*. New Delhi, India.: Asiatech Publishers Inc., pp. 85–103.
- Woodyer, R., Van der Donk, W. A. and Zhao, H. (2003) 'Relaxing the nicotinamide cofactor specificity of phosphite dehydrogenase by rational design', *Biochemistry*, 42(40), pp. 11604–11614.
- Woodyer, R., Zhao, H. and Van Der Donk, W. A. (2005) 'Mechanistic investigation of a highly active phosphite dehydrogenase mutant and its application for NADPH regeneration', *FEBS Journal*, 272(15), pp. 3816–3827.
- Wu, H., Zhang, Z., Hu, S. and Yu, J. (2012) 'On the molecular mechanism of GC content variation among eubacterial genomes.', *Biology direct*, 7(2).
- Xu, K., Ren, C., Liu, Z., Zhang, T., Zhang, T., Li, D., Wang, L., Yan, Q., Guo, L., Shen, J. and Zhang, Z. (2015) 'Efficient genome engineering in eukaryotes using Cas9 from *Streptococcus thermophilus*', *Cellular and Molecular Life Sciences*, 72(2), pp. 383–399.
- Xue, R. and Woodley, J. M. (2012) 'Process technology for multi-enzymatic reaction systems', *Bioresource Technology*. Elsevier Ltd, 115, pp. 183–195.
- Yan, B., Ding, Q., Ou, L. and Zou, Z. (2014) 'Production of glucose-6-phosphate by glucokinase coupled with an ATP regeneration system', *World Journal of Microbiology and Biotechnology*, 30(3), pp. 1123–1128.

- Yang, S., Qin, Z., Duan, X., Yan, Q. and Jiang, Z. (2015) 'Structural insights into the substrate specificity of two esterases from the thermophilic *Rhizomucor miehei*.' *Journal of lipid research*, 56(8), pp. 1616–1624.
- Yang, Z. and Pan, W. (2005) 'Ionic liquids: Green solvents for nonaqueous biocatalysis', *Enzyme and Microbial Technology*, 37(1), pp. 19–28.
- De Yin, L., Bernhardt, P., Morley, K. L., Jiang, Y., Cheeseman, J. D., Purpero, V., Schrag, J. D. and Kazlauskas, R. J. (2010) 'Switching catalysis from hydrolysis to perhydrolysis in *Pseudomonas fluorescens* esterase', *Biochemistry*, 49(9), pp. 1931–1942.
- Yokoyama, S., Matsuo, Y., Hirota, H., Kigawa, T., Shirouzu, M., Kuroda, Y., Kurumizaka, H., Kawaguchi, S., Ito, Y., Shibata, T., Kainosho, M., Nishimura, Y., Inoue, Y. and Kuramitsu, S. (2000) 'Structural genomics projects in Japan', *Progress in Biophysics and Molecular Biology*, 73(5), pp. 363–376.
- Zhang, Y.-H. P. (2015) 'Production of biofuels and biochemicals by in vitro synthetic biosystems: Opportunities and challenges', *Biotechnology Advances*, 33(7), pp. 1467–1483.
- Zhao, H. and Van Der Donk, W. A. (2003) 'Regeneration of cofactors for use in biocatalysis', *Current Opinion in Biotechnology*, 14(6), pp. 583–589.
- Zheng, M., Zhang, S., Ma, G. and Wang, P. (2011) 'Effect of molecular mobility on coupled enzymatic reactions involving cofactor regeneration using nanoparticle-attached enzymes', *Journal of Biotechnology*, 154(4), pp. 274–280.
- Zou, Y., Zhang, H., Brunzelle, J. S., Johannes, T. W., Woodyer, R., Hung, J. E., Nair, N., Van Der Donk, W. a., Zhao, H. and Nair, S. K. (2012) 'Crystal structures of phosphite dehydrogenase provide insights into nicotinamide cofactor regeneration', *Biochemistry*, 51(21), pp. 4263–4270.